Journal of General Virology (2000), **81**, 853–879. Printed in Great Britain

REVIEW ARTICLE

Virus-encoded proteinases and proteolytic processing in the *Nidovirales*

John Ziebuhr,[1] Eric J. Snijder[2] and Alexander E. Gorbalenya[3]

[1] Institute of Virology and Immunology, University of Würzburg, Versbacher Str. 7, 97078 Würzburg, Germany

[2] Department of Virology, Leiden University Medical Center, LUMC P4-26, PO Box 9600, 2300 RC Leiden, The Netherlands

[3] Advanced Biomedical Computing Center, 430 Miller Dr., Rm 235, SAIC/NCI-FCRDC, Frederick, MD 21702-1201, USA

Introduction

On the basis of similarities in their genome organization and replication strategy, RNA viruses can now be classified into 'supergroups' that often include both animal and plant viruses (Goldbach & Wellink, 1988; Strauss & Strauss, 1988). This concept is also increasingly reflected in the taxonomy of viruses; in particular by the introduction of the taxon 'order', which combines virus families for which a common ancestry seems highly probable (Mayo & Pringle, 1998). For the positive-stranded, enveloped coronaviruses and arteriviruses, which have recently been unified in the order *Nidovirales* (Cavanagh, 1997), a close phylogenetic relationship has been established on the basis of their similar polycistronic genome organization, the use of common transcriptional and (post)-translational strategies and the conservation of an array of homologous replicase domains (den Boon *et al.*, 1991). Thus, it is possible to draw a common outline of the nidovirus life-cycle (Fig. 1) (for reviews see Lai & Cavanagh, 1997; de Vries *et al.*, 1997; Snijder & Meulenberg, 1998). In some respects, however, the two virus families differ significantly from each other. For example, the largest coronavirus genome, that of mouse hepatitis virus (MHV) with 31·5 kb, is about two and a half times the size of the smallest arterivirus genome, the 12·7 kb RNA of equine arteritis virus (EAV). Also, the structural proteins of both virus families are not evidently related, resulting in important differences in virion size and structure (den Boon *et al.*, 1991; Snijder & Spaan, 1995; de Vries *et al.*, 1997).

Most major groups of positive-stranded RNA viruses of animals produce either a single polyprotein or separate non-structural and structural precursor polypeptides that are subsequently cleaved by virus-encoded or host-encoded proteinases to produce functional subunits (Dougherty & Semler, 1993). In contrast, the nidovirus structural proteins, which are encoded in the 3′-proximal region of the genome, are individually expressed from a nested set of subgenomic mRNAs (Fig. 1) generated by a unique discontinuous transcription mechanism (Spaan *et al.*, 1983; Lai *et al.*, 1984; van Marle *et al.*, 1999*a*). Apparently, virus-encoded proteinases are not involved in the maturation of the structural proteins; although a number of nidovirus envelope proteins undergo proteolytic processing by cellular proteinases as they follow the exocytotic pathway. Thus, for example, signal sequences are co-translationally removed and the coronavirus spike glycoprotein is subject, in some cases, to a host-directed maturation cleavage (Cavanagh, 1995).

In contrast to the expression of the structural genes, nidovirus-encoded proteinases play a prominent role in the expression of the replicase gene. The replicase proteins are encoded by two large, 5′-proximal open reading frames (ORFs) that occupy approximately two-thirds to three-quarters of the genome (Fig. 1). The division of the replicase gene into ORFs 1a and 1b, which are connected by a ribosomal frameshift site (Brierley *et al.*, 1987), is one of the hallmarks of the nidoviruses. It results in the translation of an ORF1a protein and a carboxyl-extended ORF1ab frameshift protein; also known as replicase polyproteins 1a and 1ab (pp1a and pp1ab)†. The size of the frameshift protein ranges from 3175 amino acids for the arterivirus EAV to about 7200 amino acids for the coronavirus MHV. The nidovirus ORF1a and ORF1ab translation products are polyprotein precursors which are cleaved by viral proteinases at a minimum of 10 (arteriviruses) or 13 (coronaviruses) sites.

Comparative sequence analyses (Boursnell *et al.*, 1987; Gorbalenya *et al.*, 1989*b*; Bredenbeek *et al.*, 1990*a*; Snijder *et al.*, 1990; den Boon *et al.*, 1991; Lee *et al.*, 1991; Godeny *et al.*, 1993; Herold *et al.*, 1993; Meulenberg *et al.*, 1993; Eleouet *et al.*, 1995) and recent experimental data (van Dinten *et al.*, 1997,

Author for correspondence: John Ziebuhr.
Fax +49 931 2013934. e-mail ziebuhr@vim.uni-wuerzburg.de

† It should be noted that the names ORF1ab polyprotein and replicase polyprotein 1ab do not refer to a corresponding ORF1ab in the viral genome, which, clearly, does not exist. The combination of the letters a and b is an abbreviation, indicating the fact that this protein is expressed from two ORFs, ORF1a and ORF1b.

0001-6806

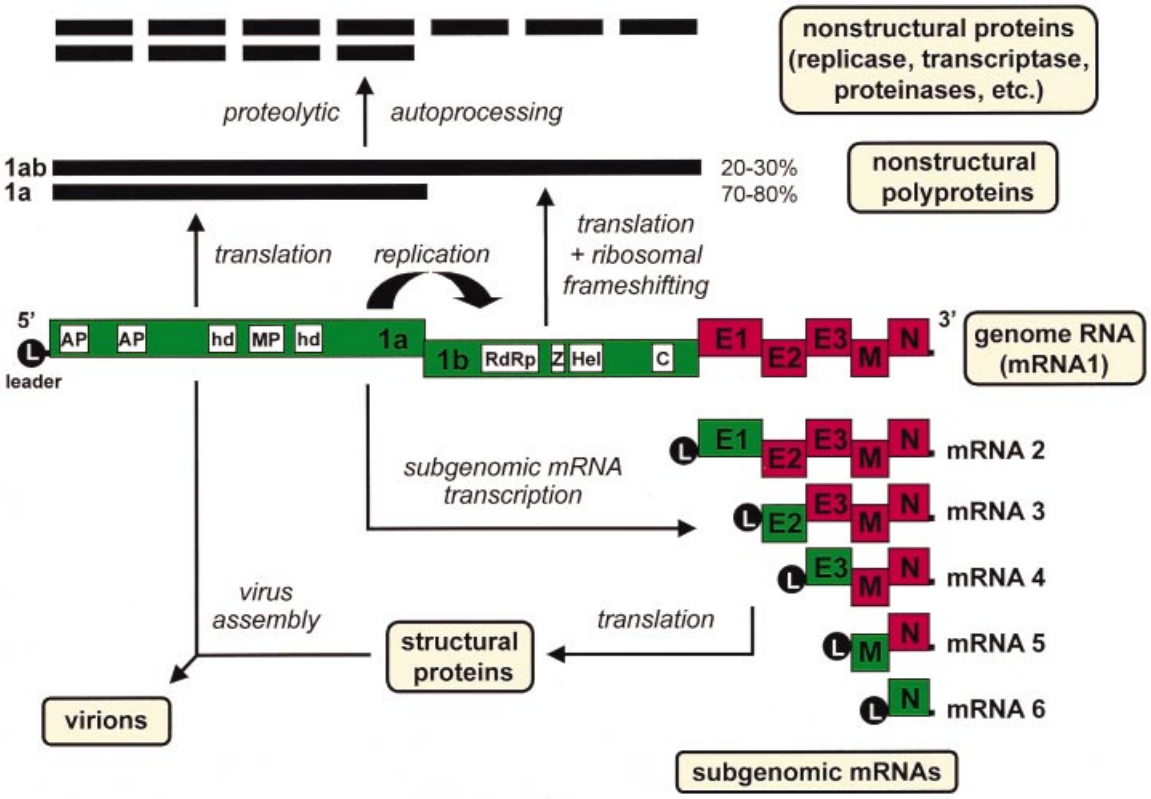

Fig. 1. Outline of the nidovirus life-cycle showing the most important similarities between coronaviruses and arteriviruses. ORFs in the polycistronic genome are indicated as boxes. ORFs that are translated from specific mRNAs are shown in green, while the downstream, non-translated ORFs are shown in red. The replicase gene, encompassing ORFs 1a and 1b, envelope protein genes (E1 to E3) that may vary in number and the genes for the triple-spanning membrane (M) protein and nucleocapsid (N) protein are shown. It should be stressed that this outline is a generalization. For example, the M and N genes are usually, but not always, the two most 3′-proximal ORFs in the genome. Many nidovirus genomes contain a variable number of additional (structural and non-structural) genes and, consequently, more than five subgenomic mRNAs are often produced. Furthermore, subgenomic mRNA transcription and genome replication involve the synthesis of minus-stranded intermediates that are not depicted here. The circle (L) at the 5′ end of the genome represents the common leader sequence that is also present at the 5′ end of the subgenomic mRNAs that are shown below the genome. The two replicase polyproteins (1a and 1ab) are depicted as solid black lines. The processing products are depicted as interrupted lines. The conserved domains/functions encoded by the replicase gene are abbreviated as follows: AP, accessory proteinase; hd, hydrophobic domain; MP, main proteinase; RdRp, RNA-dependent RNA polymerase; Z, putative zinc finger; HEL, NTPase/RNA helicase; C, conserved domain specific for nidoviruses. The amino-terminal parts of the nidovirus replicase polyproteins are processed by accessory proteinases, whereas the remainders of the polyproteins are cleaved by the main proteinase.

1999) suggest that several ORF1b-encoded replicase subunits are directly involved in viral RNA synthesis. The translational downregulation of one of these functions, the (putative) viral RNA-dependent RNA polymerase (RdRp), can also be found in a number of other virus systems of which the alphaviruses have been characterized in most detail (for a review, see Strauss & Strauss, 1994).

With respect to the ORF1a-encoded subunits of the nidovirus replicase, two main functions have emerged so far. First, hydrophobic domains in the arterivirus ORF1a protein have been shown to mediate the membrane association of the replication complex and to be able to dramatically alter the architecture of host cell membranes (van der Meer *et al.*, 1998; Pedersen *et al.*, 1999). A similar role can also be expected for the corresponding hydrophobic segments of the coronavirus replicase polyproteins (Shi *et al.*, 1999; van der Meer *et al.*, 1999). Second, the ORF1a-encoded regions of the nidovirus replicase polyproteins harbour a variety of proteolytic activities, which will be the topic of this review. Although our knowledge of the biochemical and structural properties of the nidovirus proteinases is still very limited, the available data underline the idea that, as for many other positive-stranded RNA viruses, these enzymes fulfil a crucial role in the regulation of the virus life-cycle.

This review article is organized into five main sections. In the first section, we introduce nidovirus proteinases and classify them into main and accessory proteinases. In the next two sections, we present a brief overview of the two classes of the nidovirus proteinases and, in this context, the coronavirus and arterivirus enzymes are compared to each other and to the prototypic proteinases. This is followed by a detailed description of the nidovirus proteinases themselves. The article is

Table 1. Classification of nidovirus proteinases

Virus family (virus)	Proteinase	Alternative names	Associate end-protein	Principal nucleophile	Predicted fold (alternative names)	Substrate specificity, P1 | P1′	RNA virus prototype proteinase	Cellular prototype proteinase	Classification*	Functional class
Arterivirus	PCP1α		nsp1α; nsp1 in EAV	Cys (replaced by Lys in EAV)†	$\alpha+\beta$ (papain-like)	NK	FMDV Lpro	Papain	Clan CC; Family C31; MEROPS ID, C31.001	Accessory
Arterivirus	PCP1β	PCP for EAV	nsp1β; nsp1 in EAV	Cys	$\alpha+\beta$ (papain-like)	Y,G | G	FMDV Lpro	Papain	Clan CC; Family C32; MEROPS ID, C32.001	Accessory
Arterivirus	CP2	CP, nsp2 proteinase	nsp2	Cys	NK	G | G	NK	NK	Clan CC; Family C33; MEROPS ID, C33.001	Accessory
Arterivirus	3CLSP	nsp4, SP	nsp4	Ser	Two-β-barrel (chymotrypsin-like)	E | S,G,K Q | S (3C-like)	Picornavirus 3C	Chymotrypsin	Clan SA; Family S32; MEROPS ID, S32.001	Main
Coronavirus (MHV, HCoV)	PL1^{pro}	L-pro, PLP1, PCP1	NK	Cys	$\alpha+\beta$ (papain-like)	G | V,N A | G	FMDV Lpro	Papain	Clan CC; Family C16; MEROPS ID, C16.001	Accessory
Coronavirus (IBV)	PL^{pro}	M-pro, PLP, PCP	NK	Cys	$\alpha+\beta$ (papain-like)	G | G	FMDV Lpro	Papain	Clan CC; Family C29; MEROPS ID, C29.001	Accessory
Coronavirus (MHV, HCoV)	PL2^{pro}	M-pro, PLP2, PCP2	NK	Cys	$\alpha+\beta$ (papain-like)	NK	FMDV Lpro	Papain	Clan CC; Family C29; MEROPS ID, C29.001	Accessory
Coronavirus	3CL^{pro}		3CL^{pro}	Cys	Two-β-barrel (chymotrypsin-like)	Q | S,G,A,C,N (3C-like)	Picornavirus 3C	Chymotrypsin	Clan CB; Family C30; MEROPS ID, C30.001	Main

* Gorbalenya *et al.* (1998*a*, *b*), Snijder *et al.* (1998*a*, *b*), Schiller & Baker (1998), Gibson & Denison (1998).
† The Cys-to-Lys replacement inactivates the proteolytic activity of EAV PCP1α (see text for details).
NK, Not known.

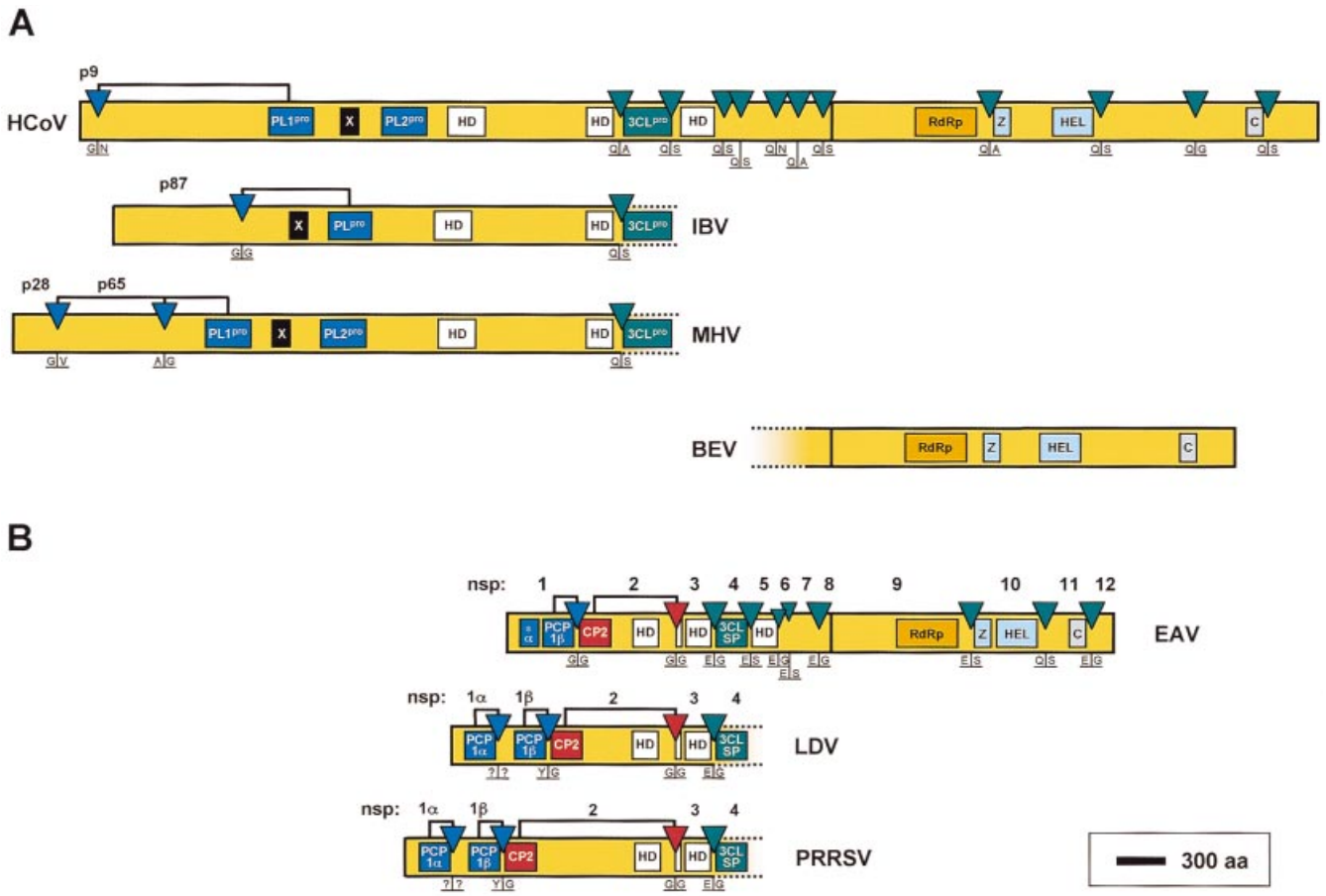

Fig. 2. Overview of the proteolytic processing of coronavirus and arterivirus replicases. The HCoV (A) and EAV (B) replicase ORF1ab polyproteins are shown with their three proteinase domains, corresponding cleavage sites (P1 and P1′ residues indicated), a number of conserved domains and the cleavage product nomenclature used in this paper. Since the processing end-products of the EAV replicase are completely known, they can be successively numbered from nsp1 to nsp12. For coronaviruses, an equivalent nomenclature cannot be given since the mapping of cleavage sites is still incomplete (see text for details). The diagrams display the differences found in the amino-terminal part of the IBV/MHV (A) or LDV/PRRSV (B) ORF1a-encoded regions of the replicase polyproteins. X, A domain with an unknown function in coronaviruses that is also conserved in alphaviruses, rubiviruses and hepatitis E virus. The α* domain in EAV represents the inactivated PCP1α domain. Below the HCoV ORF1b-encoded portion of the replicase polyprotein, the corresponding part of the replicase polyprotein of the torovirus Berne virus (BEV) is depicted, for which cleavage sites have not yet been identified. A black line indicates the position of the ribosomal frameshift site that separates ORF1a and ORF1b. Hydrophobic domains are indicated with HD. For other abbreviations, see text and the legend to Fig. 1.

concluded by two sections that describe the regulatory role of proteinases during virus replication and then give an outline of future perspectives concerning nidovirus proteinases. The reader will note that the *in vitro* characterization of the proteinases of coronaviruses is more advanced than that of arteriviruses. In contrast, current knowledge on nidovirus proteolytic regulation *in vivo* is essentially derived from research on arteriviruses.

'Main' and 'accessory' proteinases of nidoviruses

A discussion of the diverse group of proteolytic enzymes found in the *Nidovirales* requires the introduction of a standardized nomenclature. A consensus on this matter has not yet been reached. Indeed, different names are used to describe the same proteinase (Table 1). In most cases, these names allude to the relationships between the nidovirus proteinases and other proteolytic enzymes of viral or cellular origin, the so-called prototypic proteinases. However, they may also refer to major structural or biochemical properties of the proteinase (e.g. fold, catalytic system or substrate specificity; Table 1) or to its position in the virus-encoded polyprotein (a characteristic feature of positive-stranded RNA virus proteinases). Thus, the current names can become both complex and confusing. This is particularly true if combinations of different, sometimes even incompatible, properties are used, or if the specific property to which the name refers is not clear. For example, the designation 'chymotrypsin-like (CHL) proteinase' is meaningless, unless it is clearly stated whether this name refers to (i) the catalytic

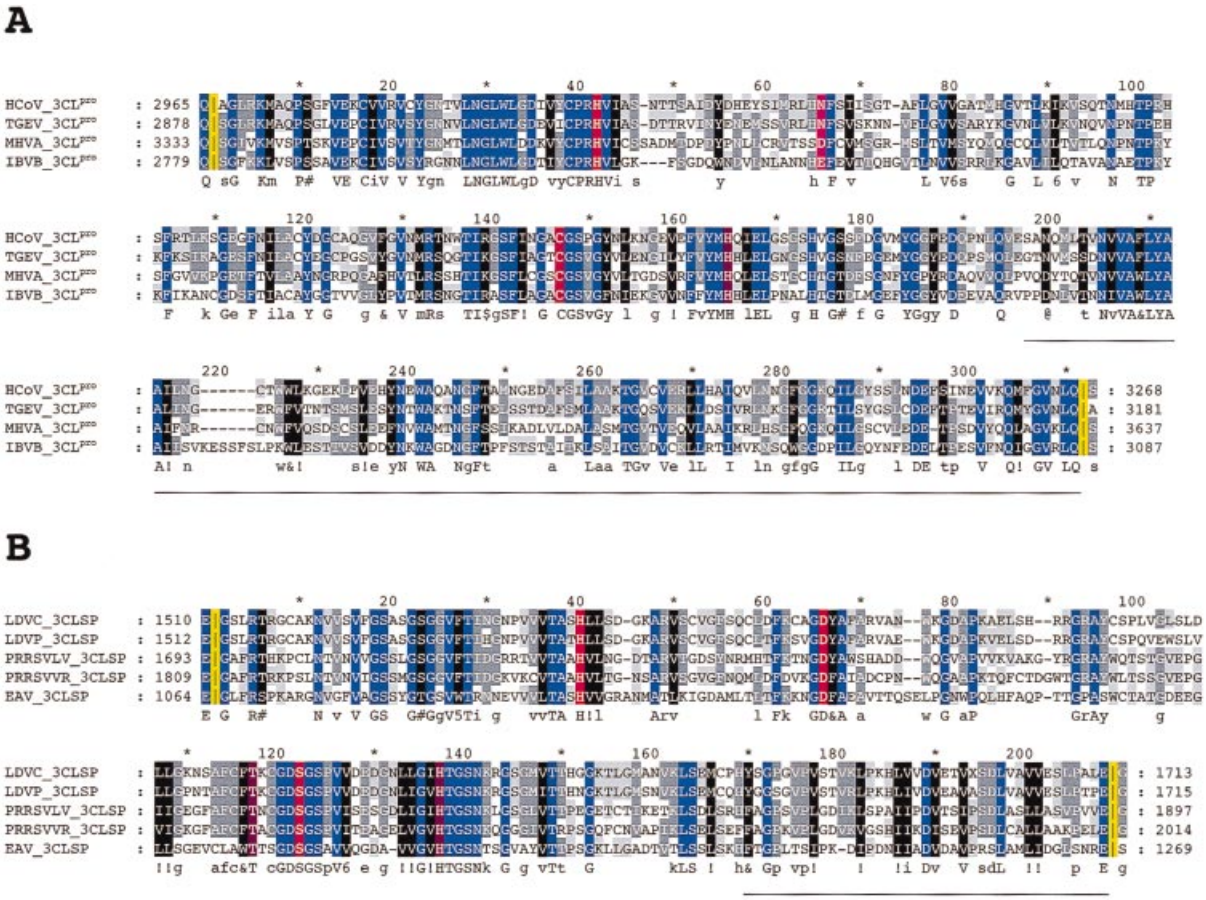

Fig. 3. Multiple alignments of nidovirus main proteinases. The alignments of coronavirus 3CL^{pro} domains (A) and arterivirus 3CLSPs (B) were generated using the ClustalX program (Thompson *et al.*, 1997). The positions of the amino-terminal and carboxyl-terminal residues of the alignment are indicated at the left and right sides, respectively, and follow the ORF1a polyprotein numbering. Coloured background: red, (putative) catalytic residues; pink, residues thought to be equivalent to the catalytic acidic residue (see text); violet, (putative) substrate-binding residues; blue, invariant residues; black, dark and light grey, residues conserved in 100%, 75% or 50%, respectively, of the sequences. Conservation groups: I, V, L, M; F, Y; K, R; D, N; E, Q; S, T. Yellow colouring, along with a vertical line marks the position of cleavage sites. The carboxyl-terminal domain (extension) is underlined but its border with the amino-terminal enzymatic domain is provisional. LDVC and LDVP, lactate dehydrogenase-elevating virus neurovirulent type C (Godeny *et al.*, 1993) and strain Plagemann (Palmer *et al.*, 1995), respectively; PRRSVLV and PRRSVVR, porcine reproductive and respiratory syndrome virus strain Lelystad (Meulenberg *et al.*, 1993) and strain ATCC VR-2332 (Nelsen *et al.*, 1999), respectively; MHVA, mouse hepatitis virus strain A59 (Bonilla *et al.*, 1994); IBVB, avian infectious bronchitis virus strain Beaudette (Boursnell *et al.*, 1987); for the other viruses see text and Tables. The National Center for Biological Information sequence ID: EAV, 133455; LDVC, 293038; LDVP, 564003; PRRSVLV, 482963; PRRSVVR, 4580009; IBVB, 138147; HCoV, 464694; TGEV, 872319; MHVA, 453423 (nucleotide).

system, (ii) the substrate specificity, (iii) the fold of the proteinase, or (iv) to an additional property in which the enzyme resembles chymotrypsin. Throughout this review, we will use the names given in the second column of Table 1. Depending on the topic discussed, these names may further be elaborated to underscore a specified property.

In this review, we will discuss the nidovirus proteinases in the context of their role as either 'main' or 'accessory' (leader) proteinases (Gorbalenya *et al.*, 1991). The main proteinase is defined as the enzyme that directly mediates the expression of the two most conserved replicase domains, RdRp and RNA helicase (HEL). We believe that this classification can be universally applied to the proteinases of all positive-stranded RNA virus families and, furthermore, it will facilitate their comparison and characterization.

Nidovirus main proteinases

Overview

All arteriviruses and coronaviruses encode one main proteinase. For the reasons specified below, this main proteinase is commonly referred to as '3C-like' (after the 3C proteinases of the *Picornaviridae*‡). Thus, the coronavirus

‡ The name 3C indicates that the associated proteinase occupies a position between two other domains, 3B and 3D, in the P3 region of the picornavirus polyprotein (Rueckert & Wimmer, 1984).

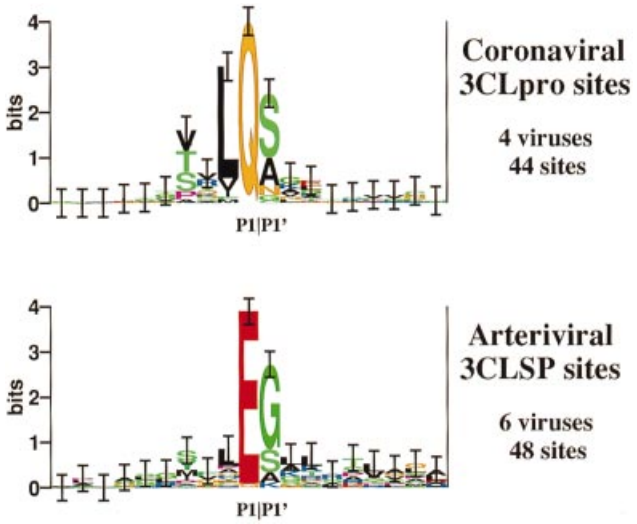

Fig. 4. Conservation of sites cleaved by nidovirus main proteinases. Two separate multiple, gap-free alignments around the P1|P1′ positions of the sites cleaved, or predicted to be cleaved, by arterivirus 3CLSPs and coronavirus 3CL^{pro} domains, respectively, were converted to logos presentations (Schneider & Stephens, 1990) in which the size of an amino acid is proportional to its conservation at the specific position and the sampling size. The amino acid conservation is measured in bits of information plotted on a vertical axis whose upper limit is determined by the natural diversity of amino acids (twenty) expressed as a logarithm of two. One standard deviation of the information content at each position is indicated by vertical bars.

enzyme will be called coronavirus 3C-like proteinase, 3CL^{pro}. Since the nucleophilic, catalytic residue of the arterivirus main proteinase is a serine and differs from that of other 3C-like proteinases, i.e. a cysteine, we have decided to stress this specific property by using the designation 3C-like serine proteinase, 3CLSP, for the arterivirus enzyme.

The 3C-like proteinases of arteriviruses and coronaviruses occupy comparable positions in the replicase polyproteins (Fig. 2). They reside upstream of the ribosomal frameshift site and domains, including RdRp and HEL, which belong to the most conserved domains in this virus order. The 3C-like proteinases are autocatalytically processed at flanking sites§ (Fig. 3A, B) and direct the proteolytic processing of all downstream domains of the replicase polyproteins, in both cases at similarly positioned sites (Fig. 2). This central role in the expression of the major replicative proteins justifies the designation of the 3C-like proteinases as the 'main' proteinase of nidoviruses.

Nidovirus 3C-like proteinases and the prototypic 3C proteinases of the *Picornaviridae* possess similar P1|P1′ substrate specificity (Fig. 4), which is determined, in part, by the conserved substrate-pocket His residue. They are also likely to employ similar folds and because of these similarities the nidovirus proteinases are '3C-like'. However, nidovirus 3C-like enzymes differ in at least two significant respects from the picornavirus 3C proteinases and their relatives, the 3C-like proteinases of the picornavirus-like supergroup (see also Ryan & Flint, 1997). First, the nidovirus 3C-like proteinases use catalytic residues that are similar but not identical to the triad His, Asp (Glu) and Cys (hereafter, the linear sequence order is given) found in 3C/3C-like proteinases of the picornavirus-like supergroup. Second, both the nidovirus and the picornavirus-supergroup 3C/3C-like proteinases belong to (and process) a highly conserved array of replicative domains whose linear orders are conserved and supergroup (order)-specific. In the picornavirus-like supergroup, this conserved arrangement of replicative domains can be described as HEL-3C(L)-RdRp, which differs from the elaborated formula of the *Nidovirales*: PL^{pro}-HD1–3CL-HD2-RdRp-Zn-HEL-C; where 3C(L) is 3C/3C-like proteinase, PL^{pro} is papain-like proteinase, HD1 and HD2 are hydrophobic domains 1 and 2, Zn is a putative zinc-binding domain, and C is a conserved, nidovirus-specific domain.

The catalytic residues of 3C and 3C-like proteinases are grafted upon a two-β-barrel structure consisting of 12 antiparallel β-strands (Bazan & Fletterick, 1988; Gorbalenya *et al.*, 1989*a*; Allaire *et al.*, 1994; Matthews *et al.*, 1994; Mosimann *et al.*, 1997). This structure was originally identified in a subset of cellular serine proteinases of which chymotrypsin (or trypsin) is the prototype. Therefore, nidovirus 3C-like proteinases may also be called 'chymotrypsin-like'. In 3C and 3C-like proteinases, Cys replaces the nucleophilic Ser and, in a subset of viruses, Glu replaces the Asp of the catalytic triad found in cellular proteinases (Bazan & Fletterick, 1988; Gorbalenya *et al.*, 1989*a*; Matthews *et al.*, 1994). The 3C-like proteinases of both corona- and arteriviruses deviate from the prototypic 3C enzymes, but, surprisingly, each of them in a different direction. Whilst the arterivirus 3CLSP employs the canonical His-Asp-Ser triad that is usually found in cellular proteinases, the coronavirus 3CL^{pro} uses Cys as its catalytic nucleophile. However, the coronavirus 3CL^{pro} seems to lack a conserved acidic residue that would be equivalent to the catalytic Asp (Glu) of 3C proteinases. Furthermore, coronavirus and arterivirus 3C proteinases display a very low overall sequence similarity. Basically, the similarity is limited to the regions of active-site residues and, even there, it is barely

§ Autocatalytic cleavage can occur *in cis* or *in trans*. Technically, the nidovirus proteinases were characterized in two types of assays, with the proteinase and its substrate residing in the same molecule, or, alternatively, in different molecules. The first assay type can be called a 'monomolecular reaction' since it permits, although not exclusively, *cis* cleavage. The second assay type is a 'bimolecular reaction' that is commonly called *trans* cleavage. If a proteinase proved to be active in both types of assay, it has been shown to cleave *in trans* but is considered likely to also cleave *in cis*. In contrast, if the proteinase is only active in the first assay type, it is thought to cleave only *in cis*. A nidovirus proteinase that is active only in the 'bimolecular reaction' assay has not yet been described. We will use the terms *cis* and *trans* cleavage as defined above, although the reader should be aware that *cis*-cleavage activity has not rigorously been proven for any of the nidovirus proteinases.

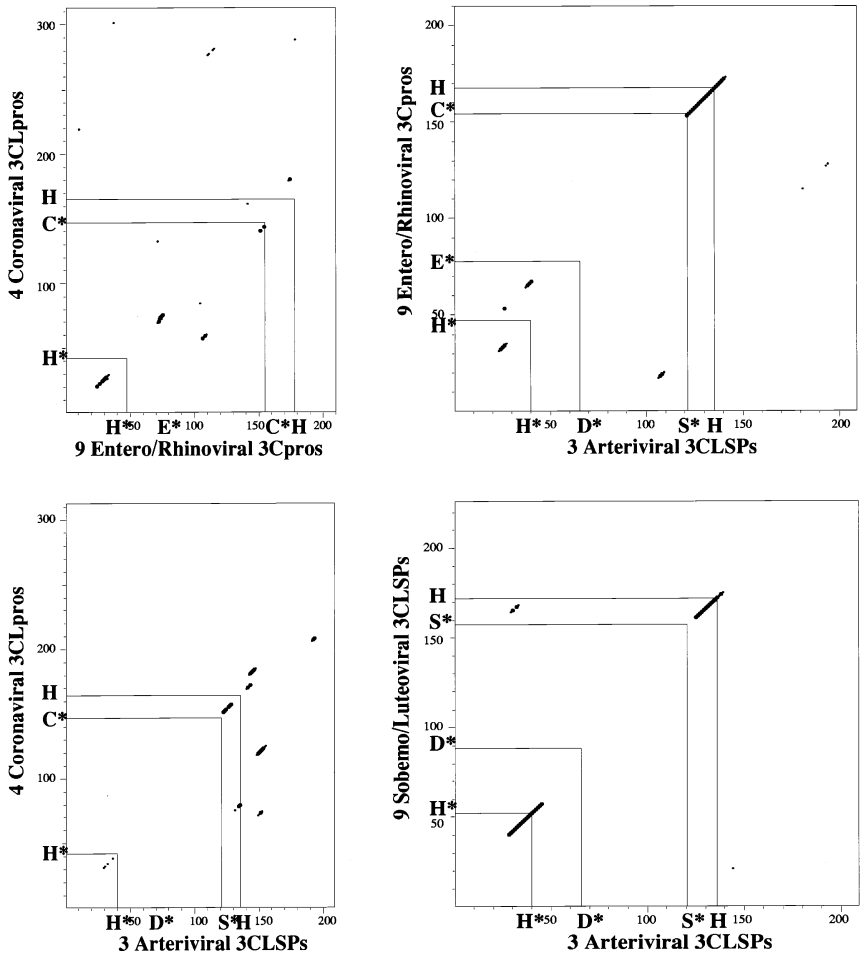

Fig. 5. Dot-plot cross-comparisons of arterivirus 3CLSPs and coronavirus 3CL^{pro} domains with other 3C/3C-like proteinases. The profiles were generated using multiple alignments of different, closely related 3C/3C-like proteinases indicated by their numbers and origins. They were compared in pairs in the dot-plot fashion using the Proplot program (Thompson *et al.*, 1994). Two profiles were compared by sliding a window of 21, 23 or 25 amino acids along each possible register and matches between two profiles that were within the top 0·5% were marked by dots. The projected positions of the catalytic residues (H*, D* or E*, C* or S*), as well as the substrate-binding H, are shown at each axis. Those dots, which lay at any of the four possible crosses of projections of two functionally equivalent residues (e.g. H* and H*) or close to a non-visible diagonal passing these crosses, belong, or may belong, to the true matches between two profiles. The rest of dots are background hits (false positives). These data are from an unpublished work of A. E. Gorbalenya.

detectable above the background in sensitive profile-versus-profile dot-plot comparisons (Fig. 5). As a result, any general alignment of the primary structures of these proteinases is largely arbitrary. The same is evident upon comparison of coronavirus 3CL^{pro} domains with the prototypic picornavirus 3C proteinases (Fig. 5). In contrast, arterivirus 3CLSPs and picornavirus 3C proteinases are clearly similar in at least one region, which includes the conserved catalytic nucleophile (Ser or Cys, respectively) and the substrate-binding His. An even more pronounced, albeit still limited, similarity can be traced between arterivirus 3CLSPs and the putative proteolytic enzymes (also called 3C-like proteinases) of sobemoviruses and luteoviruses; two families of positive-stranded RNA plant viruses (Gorbalenya *et al.*, 1988; Snijder & Gorbalenya, 1996). This similarity is correlated with (and probably related to) the conservation of the same set of active-site residues in these enzymes. Thus, despite their similar name and function, the coronavirus and arterivirus 3C-like proteinases have diverged pronouncedly from each other and from the prototypic proteinases. While the coronavirus enzymes are likely to form a separate, highly diverged branch of the two-β-barrel proteinases, the arterivirus 3CLSPs may group together with non-nidovirus proteinases. These and other striking differences should be kept in mind when the general features of nidovirus proteinases and nidovirus replicase processing are discussed.

Main proteinase-mediated processing of the coronavirus replicase polyproteins

The existence of a nidovirus 3C-like proteinase was initially predicted through computer-assisted, comparative sequence

Table 2. End-products of the 3CL^{pro}-mediated proteolytic processing of the coronavirus replicase polyproteins pp1a and pp1ab

No.	Virus*	Amino acids (aa) in the replicase polyproteins pp1a/pp1ab†	Detection in virus-infected cells	*Putative* functional domain‡	References§
1‖	HCoV	?-Gln2965	−	*HD*1	(1, 2)
	IBV	?-Gln2779	−	*HD*1	(3, 4)
	MHV	?-Gln3333	−	*HD*1	(5, 6, 7, 8, 9, 10)
2	HCoV	Ala2966-Gln3267 (302 aa)	+	3CL^{pro}	(1, 2)
	IBV	Ser2780-Gln3086 (307 aa)	−	3CL^{pro}	(3, 4)
	MHV	Ser3334-Gln3635 (302 aa)	+	3CL^{pro}	(5, 6, 7, 8, 9, 10)
3	HCoV	Ser3268-Gln3546 (279 aa)	−	*HD*2	(2, 11, 12)
	IBV	Ser3087-Gln3379 (293 aa)	−	*HD*2	(3, 4)
	MHV	Ser3636-Gln3921 (286 aa)	−	*HD*2	(5, 13)
4	HCoV	Ser3547-Gln3629 (83 aa)	+	?	(12)
	IBV	Ala3380-Gln3462 (83 aa)	−	?	(3, 14)
	MHV	Ser3922-Gln4013 (92 aa) Ser3922-Gln4010? (89 aa)	−	?	(5, 13, 15)
5	HCoV	Ser3630-Gln3824 (195 aa)	+	?	(12)
	IBV	Ser3463-Gln3672 (210 aa)	+	?	(14)
	MHV	Ser4014-Gln4207 (194 aa)	+	?	(15)
6	HCoV	Asn3825-Gln3933 (109 aa)	+	?	(12)
	IBV	Asn3673-Gln3783 (111 aa)	+	?	(16)
	MHV	Asn4208-Gln4317 (110 aa)	−	?	(5, 13, 15)
7	HCoV	Ala3934-Gln4068 (135 aa)	+	*GFL*	(12, 17)
	IBV	Ser3784-Gln3928 (145 aa)	−	*GFL*	(3, 16, 18, 19)
	MHV	Ala4318-Gln4454 (137 aa)	−	*GFL*	(5, 13)
8	HCoV	Ser4069-Asp4085 of pp1a (17 aa)	−	?	(17)
	IBV	Ser3929-Gly3951 of pp1a (23 aa)	−	?	(3, 19)
	MHV	Ser4455-Val4468 of pp1a (14 aa)	−	?	(5, 13)
9	HCoV	Ser4069-Gln4995 (927 aa)	+	*RdRp*	(17)
	IBV	Ser3929-Gln4868 (940 aa)	+	*RdRp*	(3, 18, 19)
	MHV	Ser4455-Gln5382 (928 aa)	−	*RdRp*	(5, 13, 20, 21)
10	HCoV	Ala4996-Gln5592 (597 aa)	+	*MB*, NTPase, *HEL*	(17, 22)
	IBV	Ser4869-Gln5468 (600 aa)	−	*MB, NTPase, HEL*	(3, 19)
	MHV	Ser5383-Gln5982 (600 aa)	+	MB, *NTPase, HEL*	(5, 13, 20, 21, 23, 24)
11	HCoV	Ser5593-Gln6110 (518 aa)	−	?	(22, 25)
	IBV	Gly5469-Gln5989 (521 aa)	−	?	(3, 26)
	MHV	Cys5983-Gln6503 (521 aa)	−	?	(5, 13, 20)
12	HCoV	Gly6111-Gln6458 (348 aa)	+	?	(25)
	IBV	Ser5990-Gln6327 (338 aa)	+	?	(3, 26)
	MHV	Ser6504-Gln6877 (374 aa)	−	?	(5, 13, 20)
13	HCoV	Ser6459-Lys6758 (300 aa)	−	?	(25)
	IBV	Ser6328-Met6629 (302 aa)	+	?	(3, 26)
	MHV	Ala6878-Lys7176 (299 aa)	+	?	(5, 13, 20, 27)

* HCoV, Human coronavirus (strain 229E; Herold *et al.*, 1993); IBV, avian infectious bronchitis virus (strain Beaudette; Boursnell *et al.*, 1987); MHV, mouse hepatitis virus (strain A59; Bredenbeek *et al.*, 1990*a*; Bonilla *et al.*, 1994).

† The amino acid numbering of the ORF1a and ORF1a/1b gene products, pp1a and pp1ab, is based upon predictions on the coronavirus −1 frameshift site (Brierley *et al.*, 1992) that is believed to occur at the 'slippery sequence', UUUAAAC. These predictions are supported by experiments on an analogous HTLV-I UUUAAAC slippery sequence in which direct amino acid sequencing of the transframe protein localized the site of the frameshift to the asparagine codon, AAC (Nam *et al.*, 1993).

‡ Putative functions are designated by the use of italics; HD1, hydrophobic domain 1; 3CL^{pro}, 3C-like proteinase; HD2, hydrophobic domain 2; GFL, growth factor-like domain; RdRp, RNA-dependent RNA polymerase; MB, metal ion-binding domain; NTPase, NTPase domain; HEL, helicase domain; ?, function unknown.

§ (1) Ziebuhr *et al.*, 1995; (2) Ziebuhr *et al.*, 1997; (3) Gorbalenya *et al.*, 1989*b*; (4) Tibbles *et al.*, 1996; (5) Lee *et al.*, 1991; (6) Lu *et al.*, 1995; (7) Lu *et al.*, 1996; (8) Lu & Denison, 1997; (9) Piñón *et al.*, 1997; (10) Schiller *et al.*, 1998; (11) Ziebuhr *et al.*, 1998; (12) Ziebuhr & Siddell, 1999; (13) Bonilla *et al.*, 1994; (14) Ng & Liu, 1998; (15) Lu *et al.*, 1998; (16) Liu *et al.*, 1997; (17) Grötzinger *et al.*, 1996; (18) Liu *et al.*, 1994; (19) Liu & Brown, 1995; (20) Bredenbeek *et al.*, 1990*a*; (21) Piñón *et al.*, 1999; (22) Heusipp *et al.*, 1997*b*; (23) Yoo *et al.*, 1995; (24) Denison *et al.*, 1999; (25) Heusipp *et al.*, 1997*a*; (26) Liu *et al.*, 1998; (27) Bredenbeek *et al.*, 1990*b*.

‖ The size of this protein and the proteolytic activity responsible for the generation of its amino terminus have not yet been identified.

analysis of the complete replicase polyprotein of an avian coronavirus, infectious bronchitis virus (IBV), and the picornavirus 3C proteinases (Gorbalenya *et al.*, 1989*b*). In the same study, the processing of a dozen putative cleavage sites by the IBV 3CL^{pro} was predicted to yield the mature components of the replicase complex. All of these sites contained Q|S,G dipeptides, which are typical substrates of 3C/3C-like proteinases. Also, the putative cleavage sites were found to have bulky hydrophobic residues at their P2 positions. These predictions were subsequently extended by sequence analyses of the replicase genes of three other coronaviruses, MHV, human coronavirus 229E (HCoV) and porcine transmissible gastroenteritis virus (TGEV) (Lee *et al.*, 1991; Herold *et al.*, 1993; Eleouet *et al.*, 1995). Specifically, both the 3CL^{pro} domain (~300 amino acids) (Fig. 3A) and the vast majority of the predicted cleavage sites were found to be conserved in coronaviruses (Figs 4 and 3A). Collectively, these studies have produced a model of the 3CL^{pro}-mediated processing of the coronavirus replicase polyprotein(s) which forms the basis of many current experimental studies.

The first experimental evidence for a coronavirus proteinase activity encoded in the 3′-proximal ORF1a sequence was reported for IBV (Liu *et al.*, 1994). In this study, the expression of the putative RdRp domain was shown to involve a virus-encoded proteinase activity that mapped to a region previously predicted to contain a 3CL^{pro} domain (Gorbalenya *et al.*, 1989*b*). It was concluded that the proteolytic activity observed was 3CL^{pro}-mediated. Evidence supporting this hypothesis was obtained from three studies on the 3CL^{pro} domains of IBV, MHV and HCoV 229E (Liu & Brown, 1995; Lu *et al.*, 1995; Ziebuhr *et al.*, 1995). It is worth noting that, in these initial studies, the coronavirus 3CL^{pro} proved to be active in quite different expression systems. The MHV 3CL^{pro} was expressed *in vitro* in a rabbit reticulocyte lysate, the HCoV 229E 3CL^{pro} was expressed in *Escherichia coli*, and the IBV 3CL^{pro} was expressed in Vero cells using the recombinant vaccinia virus/T7 system. To further analyse the 3CL^{pro}-mediated processing of the coronavirus replicase proteins, two of the systems mentioned above were exploited. In the case of IBV, the 3CL^{pro} was co-expressed with substrate proteins derived from different portions of the replicase polyprotein(s) and 3CL^{pro}-mediated proteolysis of substrates containing specific, mostly predicted, cleavage sites was observed (Liu *et al.*, 1994, 1997, 1998; Liu & Brown, 1995; Ng & Liu, 1998). The cleavage products were identified by their apparent molecular mass in SDS–PAGE, and the cleavage sites were mapped by site-directed mutagenesis. For HCoV and MHV, the assay systems used were based upon bacterially expressed 3CL^{pro} domains (Ziebuhr *et al.*, 1995; Herold *et al.*, 1996; Seybert *et al.*, 1997). This approach greatly facilitated the identification and N-terminal sequence analysis of 3CL^{pro} cleavage sites by using both *in vitro*-translated and recombinant protein substrates (Ziebuhr *et al.*, 1995; Grötzinger *et al.*, 1996). Furthermore, it allowed for the determination of kinetic parameters using synthetic peptides combined with quantitative analysis of the substrate conversion (Ziebuhr & Siddell, 1999). Finally, in the MHV system, it has even proved possible to perform amino-terminal microsequence analysis of metabolically labelled 3CL^{pro} cleavage products isolated by immunoprecipitation from virus-infected cells (Lu *et al.*, 1998).

The information collected during studies with different coronaviruses can now be used to map the 3CL^{pro} processing sites in the coronavirus replicase polyprotein (Fig. 2A). Taken together, at least 12 processing end-products (including the 3CL^{pro} itself) are generated by 3CL^{pro}-mediated cleavage. The processing products of HCoV, IBV and MHV that have been identified so far are summarized in Table 2. Although the proteolytic processing of TGEV, another coronavirus with a known genome sequence (Eleouet *et al.*, 1995), has not yet been characterized experimentally, its close relationship with HCoV allows for reliable functional predictions.

Catalytic centre and substrate specificity of the coronavirus main proteinase

Based on the analysis of the IBV sequence (Boursnell *et al.*, 1987), the catalytic system of the coronavirus 3CL^{pro} was proposed to resemble that of other viral 3C and 3C-like proteinases (Gorbalenya *et al.*, 1989*b*) and to involve a His-Asp(Glu)-Cys catalytic triad. In accordance with these predictions, mutagenesis studies with three different coronavirus 3CL^{pro} domains provided experimental evidence that, for HCoV, His-3006 and Cys-3109, and the equivalent residues in IBV and MHV (Fig. 3A), are indispensable for activity (Liu & Brown, 1995; Lu *et al.*, 1995; Ziebuhr *et al.*, 1995, 1997; Tibbles *et al.*, 1996; Seybert *et al.*, 1997). Interestingly, the replacement of the putative active-site Cys by Ser in the IBV 3CL^{pro} was reported to produce an enzyme with residual activity in a *cis*-cleavage assay, supporting the relatedness of the coronavirus 3CL^{pro} with serine proteinases of the chymotrypsin family (Tibbles *et al.*, 1996). In contrast with this finding, no activity was detected for the corresponding Cys-to-Ser mutants of the MHV and HCoV 3CL^{pro} domains using *trans*-cleavage assays (Seybert *et al.*, 1997; Ziebuhr *et al.*, 1997). The reasons for this discrepancy remain to be investigated, but may reside in the different sensitivities of *trans*- versus *cis*-cleavage assays. In conclusion, the available data strongly suggest that the conserved His and Cys residues represent the general base and nucleophile, respectively, of a charge-relay system similar to that of serine proteinases.

Special efforts have been made to identify the coronavirus 3CL^{pro} counterpart of the third (acidic) residue present in the catalytic triad of CHL enzymes. In the original analysis of the IBV replicase sequence, Glu-2843 was aligned with the catalytic acidic residue of 3C and 3C-like proteinases (Gorbalenya *et al.*, 1989*b*). However, subsequent sequence analyses of other coronavirus replicases revealed that Glu-2841 is more conserved than Glu-2843. Also, Glu-2841 is the only residue in

the region delimited by the catalytic His and Cys (Fig. 3A; cf. Fig. 3B) whose variability in coronaviruses (Asp in MHV, Glu in IBV, and Asn in HcoV and TGEV) could somehow be reconciled with the role of the third catalytic residue (Gorbalenya & Koonin, 1993; Gorbalenya & Snijder, 1996). Consequently, this position was probed by site-directed mutagenesis. However, variable results were obtained. For example, the substitution of Glu-2841 by Gln in IBV resulted in an active enzyme (Liu & Brown, 1995). Likewise, the replacement of Asp-3398, the IBV Glu-2841 equivalent, by either Ala or Pro did not significantly alter the activity of the MHV 3CL^{pro} (Lu & Denison, 1997). In contrast, the replacement of the equivalently positioned Asn by Gly or Pro in the HCoV 3CL^{pro} had different effects. The replacement of Asn with Gly did not change the rate of substrate conversion as compared to the wild-type enzyme in a peptide cleavage assay (Ziebuhr *et al.*, 1997). However, the replacement of Asn with Pro substantially reduced the rate of substrate conversion. It is important to note, however, that, in these experiments, the enzymatic activity was not rigorously tested (e.g. by comparison of purified wild-type and mutant 3CL^{pro} domains in quantitative assays using a range of different substrates). Therefore, it would still be premature to draw definitive conclusions concerning the role of this (conserved) residue in the function of the coronavirus 3CL^{pro}. However, keeping in mind that all other 3C/3C-like proteinases tested so far, including those of arteriviruses, only tolerated an exchange of Glu and Asp at this position, the data seem to indicate that the coronavirus 3CL^{pro} lacks a corresponding acidic, catalytic residue in its sequence. It remains to be seen whether an alternative conserved acidic (or even non-acidic) residue outside the region delimited by the catalytic His and Cys assumes the catalytic role and occupies a position equivalent in space to that of the catalytic acidic residue of other 3C/3C-like proteinases.

The coronavirus 3CL^{pro} displays additional features that clearly separate it from other virus-encoded 3C-like proteinases, including the arterivirus main proteinase. For example, it employs a novel version of the substrate-binding pocket 'core' motif, which is characteristically Gly-X-His for most other 3C/3C-like proteinases. Thus, the Gly residue of this motif (Bazan & Fletterick, 1988; Gorbalenya *et al.*, 1989*a*, *b*) is conserved in the vast majority of serine and cysteine proteinases with CHL folds and only very few proteinases tolerate substitutions with small amino acids (Ala or Cys) at this position. This conservation pattern indicates a strong selection pressure with regard to the space that this specific residue occupies. In contrast, in all coronavirus 3CL^{pro} domains studied so far, Gly appears to be replaced by Tyr (Gorbalenya *et al.*, 1989*b*; Lee *et al.*, 1991; Herold *et al.*, 1993; Eleouet *et al.*, 1995) (Fig. 3A). Given the unusual nature of this replacement and the very low level of overall similarity between the coronaviral and the other CHL proteinases (Fig. 5), additional support for this theoretical assignment is needed. The Tyr residue of the Tyr-X-His motif has not yet been probed by mutagenesis. However, the replacement of the His residue (His-3127) by Ser completely abolished the proteolytic activity of the HCoV 3CL^{pro} (Ziebuhr *et al.*, 1997). This inactivation was selective since a similar replacement of His-3136, another conserved His residue in this region, was not so detrimental (Ziebuhr *et al.*, 1997). Thus, the importance of the Tyr-X-His motif has been confirmed, implying that coronaviruses may indeed have accepted a Gly-to-Tyr replacement during evolution. It can be expected that this replacement is coupled to other substitution(s) in the active site to accommodate the bulky side chain of Tyr. The above data are also compatible with a model, originally developed and substantiated for other 3C/3C-like proteinases (Bazan & Fletterick, 1988; Gorbalenya *et al.*, 1989*a*; Allaire *et al.*, 1994; Matthews *et al.*, 1994; Mosimann *et al.*, 1997), that implicates His-3127 (and its counterparts in other coronaviruses) in the formation of hydrogen bonds to the P1 glutamine side chain of 3CL^{pro} substrates (Gorbalenya *et al.*, 1989*b*). The high degree of conservation of coronavirus cleavage sites upstream of the P1 position (Fig. 4) suggests that the substrate-binding pocket of the coronavirus 3CL^{pro} may have numerous additional contacts with its substrates. The determinants of these interactions remain to be elucidated.

The substrate specificity of 3CL^{pro} resembles that of many other 3C/3C-like proteinases (Kräusslich & Wimmer, 1988; Dougherty & Semler, 1993; Blom *et al.*, 1996) in so far as the P1 position of the substrate is exclusively occupied by Gln and small, aliphatic residues (Ser, Ala, Asn, Gly and Cys) are found at the P1′ position (Fig. 4). However, Asn and Cys are most uncommon as P1′ residues outside of the coronaviruses, although a P1′ Asn is found in rhinoviruses (Blom *et al.*, 1996) and, in a mutagenesis study, Cys proved to be a tolerable substitution in one of the encephalomyocarditis virus (EMCV) 3Cpro sites (Parks *et al.*, 1989). In four different coronaviruses, one 3CL^{pro} cleavage site consistently contains Asn at the P1′ position (Liu *et al.*, 1997; Lu *et al.*, 1998; Ziebuhr & Siddell, 1999) and, for MHV, a P1′ Cys residue was predicted for another site (Lee *et al.*, 1991). A peptide that mimicked an HCoV 3CL^{pro} site with the P1′ Asn residue was processed relatively poorly *in vitro*, which additionally points to the exceptional nature of Asn at this position (Ziebuhr & Siddell, 1999).

Upon comparison of a large number of cleavage sites, most of which have experimentally been confirmed for at least one coronavirus (Fig. 2), it is evident that in addition to P1 and P1′, the P2, P3, P4, P2′ and P3′ positions have a restricted variability (Fig. 4). Among these, the P2 and P4 positions are most conserved with bulky hydrophobic residues (mainly Leu) at P2 and Val, Thr, Ser (and Pro) at P4 being clearly favoured (Fig. 4). A similar complexity was previously described for the primary cleavage site determinants of the potyvirus 3C-like proteinase (NIa protein). The potyvirus sites could be transferred into an alien protein where they promoted selective

cleavage of the protein by the cognate 3C-like proteinase in a reaction superficially resembling the cleavage of DNA by restriction endonucleases (Carrington & Dougherty, 1988). The coronavirus 3CL^{pro} cleavage sites can be predicted to possess similar properties. The efficiency of cleavage at specific sites is likely to be determined by the exact composition of the sites, since synthetic peptides mimicking different cleavage sites were processed in competition experiments at significantly different rates by the HCoV 3CL^{pro} (Ziebuhr & Siddell, 1999). In view of these data, it seems likely that together with the accessibility of potential cleavage sites in the context of the polyprotein the properties of the cleavage sites themselves contribute significantly to the coordinated, temporal release of specific polypeptides from the replicase polyproteins. This might lead to the (irreversible) activation or inactivation of specific functions in the course of the virus life-cycle, as has been demonstrated for a number of other positive-stranded RNA viruses.

Structural aspects of the coronavirus main proteinase

The coronavirus 3CL^{pro} domains are the largest proteinases of their type. They consist of 302–307 amino acids, whereas the prototypic poliovirus 3C proteinase contains only 182 residues. This size difference is due to the presence of a unique, carboxyl-terminal region of approximately 110 amino acids which appears to be required for proteolytic activity. Thus, a large number of different carboxyl-terminally truncated versions of the HCoV 3CL^{pro} are inactive in assays using synthetic peptides (Ziebuhr *et al.*, 1997; J. Ziebuhr, unpublished data). Also, the removal of 28 carboxyl-terminal amino acids from the MHV 3CL^{pro} abolishes its activity in an *in vitro* translation system (Lu & Denison, 1997). Recently, in apparent contrast to the HCoV and MHV data it was shown that a recombinant form of the IBV 3CL^{pro} tolerated the introduction of six consecutive His residues near its carboxyl terminus without loss of activity (Tibbles *et al.*, 1999).

In the absence of a structural model for the coronavirus 3CL^{pro}, we can only speculate on the function of the carboxyl-terminal region. Obviously, several, not mutually exclusive, functions could be related to this domain, e.g. (i) maintenance of the overall folding of the enzyme, (ii) involvement in catalysis or (iii) substrate recognition, and (iv) a non-proteolytic function. It should be noted that two other groups of 3C-like proteinases, those of arteriviruses (Fig. 3B) and potyviruses (reviewed in Ryan & Flint, 1997), also have carboxyl-terminal extensions, albeit of smaller sizes. Again no specific function(s) could be attributed to these domains.

The IBV 3CL^{pro} appears to contain structural determinants that, in reticulocyte lysates, prime this proteinase for degradation by the concerted action of ubiquitin and the 26S ATP-dependent proteinase (Tibbles *et al.*, 1995). The relevance of this observation to the turnover of the coronavirus proteinase *in vivo* has not yet been studied. For another distantly related proteinase that carries a protein destruction signal, the EMCV 3Cpro (Lawson *et al.*, 1999), a correlation between the kinetics of proteinase degradation *in vitro* and *in vivo* has been reported (Lawson *et al.*, 1994). Importantly, the 3C proteinase of another picornavirus, poliovirus, was shown to be stable (Lawson *et al.*, 1999). Thus, the proteinase degradation signal is a virus-specific structural feature in picornaviruses and, possibly, in coronaviruses.

Sequence comparisons have revealed that the coronavirus 3CL^{pro} is flanked by two hydrophobic domains, HD1 and HD2 (Gorbalenya *et al.*, 1989*b*; Lee *et al.*, 1991; Herold *et al.*, 1993; Eleouet *et al.*, 1995), that are also conserved in arteriviruses (Fig. 2). Recent data from *in vitro* translation experiments have shown that microsomal membranes are required for the efficient autoproteolytic processing of the 3CL^{pro} from HD1 and HD2, most likely by assisting in the proper folding of these proteins (Tibbles *et al.*, 1996; Piñón *et al.*, 1997; Schiller *et al.*, 1998). However, after being released from the polyprotein, the 3CL^{pro} activity does not depend on membranes, or any other cofactor(s), at least for its proteolytic activity *in vitro* (Ziebuhr *et al.*, 1995). It has been suggested that HD1 and HD2 may contribute to the intracellular localization of the 3CL^{pro} itself and, possibly, of the virus replication complex in general (Gorbalenya *et al.*, 1989*b*). Recent data, obtained by using immunofluorescence and electron microscopy, strongly support this hypothesis (Heusipp *et al.*, 1997*a*; Bi *et al.*, 1998; Schiller *et al.*, 1998; Ziebuhr *et al.*, 1998; Denison *et al.*, 1999; Shi *et al.*, 1999; van der Meer *et al.*, 1999; Ziebuhr & Siddell, 1999). Specifically, it has been found that the coronavirus nucleocapsid protein, numerous replicase gene-derived proteins and newly synthesized RNA co-localize to intracellular (mainly late endosomal) membranes (van der Meer *et al.*, 1999). However, there are also reports that favour a Golgi localization for the MHV replication complexes, at least in specific cell types (Bi *et al.*, 1998; Shi *et al.*, 1999). From the combined data, it can be concluded that coronavirus replication takes place at intracellular membranes and that a large number of non-structural, replicase gene-encoded proteins contribute to the formation and function of the coronavirus replication complex.

As outlined above, the coronavirus 3CL^{pro} has a number of unique properties that remain poorly understood due to the lack of structural information about any of these enzymes. The currently available structures of three picornavirus 3C proteinases (Allaire *et al.*, 1994; Matthews *et al.*, 1994; Mosimann *et al.*, 1997) and the results of inhibitor analyses (Tibbles *et al.*, 1996; Ziebuhr *et al.*, 1997) support the classification of the coronavirus 3CL^{pro} as a two-β-barrel-fold protein. However, they are of limited use in understanding the unique features of these very distant relatives. Recently developed expression and purification systems (Ziebuhr *et al.*, 1995, 1997; Seybert *et al.*, 1997) could provide a suitable basis for the crystallization of coronavirus 3CL^{pro} domains and the elucidation of their structure.

Table 3. End-products of the 3CLSP-mediated processing of the arterivirus ORF1a/ORF1ab polyproteins

nsp*	Virus†	Amino acids (aa) in the ORF1a/ORF1ab polyproteins	*Putative* functional domain‡	References§
3‖	EAV	Gly832-Glu1064 (233 aa)	*HD*	(1, 2, 3, 4, 5, 6)
	PRRSV	Gly1463-Glu1693 (231 aa)	*HD*	(7, 8, 9)
	LDV	Gly1287-Glu1512 (226 aa)	*HD*	(6, 10)
4	EAV	Gly1065-Glu1268 (204 aa)	3CLSP	(1, 2, 3, 4, 6)
	PRRSV	Gly1694-Glu1896 (203 aa)	*3CLSP*	(7, 8, 9)
	LDV	Gly1513-Glu1708 (196 aa)	*3CLSP*	(6, 10)
5	EAV	Ser1269-Glu1430 (162 aa)	*HD*	(1, 2, 3, 4, 6)
	PRRSV	Gly1463-Glu2066 (170 aa)	*HD*	(7, 8, 9)
	LDV	Gly1709-Glu1878 (170 aa)	*HD*	(6, 10)
6	EAV	Gly1431-Glu1452 (22 aa)	?	(1, 3, 4, 6)
	PRRSV	Gly2067-Glu2082 (16 aa)	?	(7, 8, 9)
	LDV	Gly1879-Glu1894 (16 aa)	?	(6, 10)
7	EAV	Ser1453-Glu1677 (225 aa)	?	(1, 3, 4)
	PRRSV	Ser2083-Glu2351 (269 aa)	?	(7, 8, 9)
	LDV	Gly1895-Glu2161 (267 aa)	?	(6, 10)
8	EAV	Gly1678-Asn1727 (50 aa)	?	(1, 2, 3, 4)
	PRRSV	Ala2352-Cys2397 (46 aa)	?	(7, 8, 9)
	LDV	Gly2162-Cys2206 (45 aa)	?	(6, 10)
9	EAV	Gly1678-Glu2370 (693 aa)	*RdRp*	(1, 11, 13)
	PRRSV	Ala2352-Glu3036 (685 aa)	*RdRp*	(7, 8, 9)
	LDV	Gly2162-Glu2843 (682 aa)	*RdRp*	(6, 10)
10	EAV	Ser2371-Gln2837 (467 aa)	*MB, NTPase, HEL*	(1, 11, 12, 13)
	PRRSV	Gly3037-Glu3478 (442 aa)	*MB, NTPase, HEL*	(7, 8, 9)
	LDV	Lys2844-Glu3272 (429 aa)	*MB, NTPase, HEL*	(6, 10)
11	EAV	Ser2838-Glu3056 (219 aa)	?	(1, 11, 13)
	PRRSV	Gly3479-Glu3702 (224 aa)	?	(7, 8, 9)
	LDV	Gly3273-Glu3494 (222 aa)	?	(6, 10)
12	EAV	Gly3057-Val3175 (119 aa)	?	(1, 11, 13)
	PRRSV	Gly3703-Pro3854 (152 aa)	?	(7, 8, 9)
	LDV	Gly3495-Lys3616 (122 aa)	?	(6, 10)

* nsp, Non-structural protein.
† EAV, Equine arteritis virus (Bucyrus strain; den Boon *et al.*, 1991); PRRSV, porcine reproductive and respiratory syndrome virus (Lelystad strain; Meulenberg *et al.*, 1993); LDV, lactate dehydrogenase-elevating virus (LDV-P; Palmer *et al.*, 1995).
‡ Putative functions are designated by the use of italics; 3CLSP, 3C-like serine proteinase; HD, hydrophobic domain; RdRp, RNA-dependent RNA polymerase; MB, metal ion-binding domain; NTPase, NTPase domain; HEL, helicase domain; ?, function unknown.
§ (1) den Boon *et al.*, 1991; (2) Snijder *et al.*, 1994; (3) Snijder *et al.*, 1996; (4) Wassenaar *et al.*, 1997; (5) Snijder *et al.*, 1995; (6) Godeny *et al.*, 1993; (7) Meulenberg *et al.*, 1993; (8) Nelsen *et al.*, 1999; (9) Allende *et al.*, 1999; (10) Palmer *et al.*, 1995; (11) van Dinten *et al.*, 1996; (12) van Marle *et al.*, 1999*b*; (13) van Dinten *et al.*, 1999.
‖ The amino terminus of nsp3 is generated by CP2-mediated cleavage (Snijder *et al.*, 1995).

Arterivirus main proteinase

The arterivirus 3C-like serine proteinase (3CLSP) was first identified by comparative sequence analysis of the ORF1a protein of EAV, the arterivirus prototype (den Boon *et al.*, 1991). Subsequently, the 3CLSP domain was shown to reside in a 21 kDa cleavage product (nsp4; 204 residues in the case of EAV) derived from the central region of the ORF1a protein (Fig. 2B) (Snijder *et al.*, 1994). In addition to this fully cleaved product, a number of 3CLSP-containing processing intermediates were identified in EAV-infected cells, the most abundant ones being nsp3–12, nsp3–8 and nsp3–4 (Snijder *et al.*, 1994; van Dinten *et al.*, 1996). The proteolytic activity of the 3CLSP was demonstrated in the recombinant vaccinia virus/T7 expression system (Snijder *et al.*, 1996), in which studies to characterize this proteinase (e.g. site-directed mutagenesis) were also carried out. The 3CLSP was shown to mediate (at least) eight cleavages in the EAV replicase; five in the carboxyl-terminal half of the ORF1a protein and three in the ORF1b-encoded polypeptide (Snijder *et al.*, 1996; Wassenaar *et al.*, 1997; van Dinten *et al.*, 1999). These cleavage sites were identified by comparison with other arterivirus

sequences (Godeny *et al.*, 1993; Meulenberg *et al.*, 1993; Snijder *et al.*, 1996; van Dinten *et al.*, 1996) and, subsequently, confirmed by site-directed mutagenesis and expression studies (Snijder *et al.*, 1996; Wassenaar *et al.*, 1997; van Dinten *et al.*, 1999). Two sites (nsp4|5 and nsp6|7) were also confirmed by direct amino-terminal sequence analysis of cleavage products derived in an alphavirus expression system (Wassenaar *et al.*, 1997). Four of the known EAV 3CLSP cleavages (Fig. 2B) occur at Glu|Gly sites, three at Glu|Ser dipeptides, and one (nsp9|10) between Gln and Ser. All of these sites are conserved in the known arterivirus sequences (Wassenaar *et al.*, 1997; van Dinten *et al.*, 1999) although, for a number of them, Ala (and in one case Lys) is predicted to be the P1′ residue (Fig. 4). In addition to the P1|P1′ positions, some degree of conservation (although, less than in coronaviruses) is evident for the P2 and P4 positions (Fig. 4). This appears to suggest that the residues at these positions contribute to substrate recognition but experiments to verify this hypothesis have not yet been reported. The sizes of the processing end-products generated by the 3CLSP (nsp3–12) are constant among different arteriviruses (Table 3), which is in contrast to the virus-specific size heterogeneity of proteins carrying accessory proteolytic activities (nsp1–2 region; see below).

The arterivirus 3CLSP was the first experimentally characterized 3C-like proteinase using the His-Asp-Ser catalytic triad typical of cellular CHL proteinases. Replacements of residues in the predicted catalytic triad (His-1103, Asp-1129 and Ser-1184 in EAV), using site-directed mutagenesis, confirmed that these residues are indispensable for proteolysis (Snijder *et al.*, 1996). The substitution of Asp-1129 by Glu, but not the substitution of Ser-1184 by Cys [which are the replacements identified in the catalytic centres of (some) 3C/3C-like proteinases], was partially tolerated by the EAV 3CLSP. The putative substrate-binding region of this proteinase was predicted to include the conserved Thr-1179 and His-1199 residues (Fig. 3B), whose counterparts were originally recognized in viral 3C-like cysteine proteinases (Bazan & Fletterick, 1988; Gorbalenya *et al.*, 1989*a*; Allaire *et al.*, 1994; Matthews *et al.*, 1994). These residues are assumed to determine the P1 specificity of the 3CLSP for Glu (Gln). Accordingly, the EAV 3CLSP showed a pattern of tolerance to replacements at these positions that is typical of 3C/3C-like cysteine proteinases (Snijder *et al.*, 1996). Thus, the EAV 3CLSP can be considered to be the prototype of a distinct CHL proteinase subgroup that has provisionally been named '3C-like serine proteinases' (Snijder *et al.*, 1996). Similar proteinases were predicted to be encoded by plant sobemoviruses, luteoviruses and pea enation mottle virus, as well as animal astroviruses (Gorbalenya *et al.*, 1988; Snijder *et al.*, 1996, and references therein).

The ability of the EAV 3CLSP to cleave specific sites *in cis* or *in trans* has not been studied in detail. It is likely that the two sites flanking the 3CLSP domain in the replicase polyproteins (nsp3|nsp4 and nsp4|nsp5) are cleaved *in cis*. However, the EAV 3CLSP has also been shown to be active *in trans*. For example, it has been found that, in the recombinant vaccinia virus/T7 expression system, an nsp4 expression product was able to cleave the nsp9|10 and nsp10|11 sites in a separately expressed, ORF1b-encoded polyprotein (van Dinten *et al.*, 1999).

Nidovirus accessory proteinases

Overview

All arteriviruses and coronaviruses encode between one and three 'accessory' proteinases, which are very distantly related. Thus, in the course of nidovirus evolution, duplications of accessory proteinases may have occurred (Lee *et al.*, 1991; den Boon *et al.*, 1995; Snijder *et al.*, 1995). It remains unclear whether these duplications happened in the ancestral lineage or independently (and repeatedly) in different lineages. All of the 'accessory' proteinases (i) recognize one or two sites that are located in the amino-terminal half of the replicase polyproteins, (far) upstream of the major conserved domains, (ii) cleave peptides that all contain at least one small residue at the scissile bond, (iii) have a catalytic dyad consisting of Cys and a downstream His (an arrangement that is typical of cellular proteinases related to papain), and (iv) may employ variants of the $\alpha+\beta$ fold that is conserved in this class of proteinases. For the last two reasons, the nidovirus accessory proteinases are often called 'papain-like'. Some of the conserved properties mentioned above are also reflected in the names of the accessory proteinases (Table 1). However, there are notable differences between the arterivirus and coronavirus enzymes. The catalytic Cys and His residues of arterivirus accessory proteinases are separated by rather short regions whose sizes are within a range typical of many other viral papain-like proteinases. In coronaviruses, however, these regions are almost twice as long, making the coronavirus proteinases the largest in this class of RNA virus proteins. These proteinases are separated by a region of not less than $\sim$ 1000 amino acid residues. In contrast, arterivirus accessory proteinases occupy a much more amino-proximal position in the replicase polyproteins. Also, the arterivirus enzymes cleave downstream of the catalytic domain, whereas the characterized coronavirus proteinases cleave upstream. In this review article, two similar but not identical nomenclatures will be used to designate the arterivirus and coronavirus accessory proteinases. Each of these nomenclatures is specifically designed to accommodate the unique features of specific sets of proteinases.

The arteriviruses EAV, lactate dehydrogenase-elevating virus (LDV) and porcine reproductive and respiratory syndrome virus (PRRSV) encode a cassette of three adjacent proteolytic domains that are known (from amino terminus to carboxyl terminus) as papain-like cysteine proteinase 1α (PCP1α), papain-like cysteine proteinase 1β (PCP1β) and cysteine proteinase of nsp2 (CP2). The digits in the names stand for the non-structural proteins (nsp; numbered from

amino to carboxyl terminus) in which the proteolytic domains reside. PCP1α and PCP1β may reside in one protein (nsp1; EAV) or in two separate proteins (nsp1α and nsp1β; LDV and PRRSV) that precede nsp2, which itself contains CP2. The EAV PCP1α domain is an inactivated enzyme and, therefore, the cleavage between the PCP1α and PCP1β domains does not occur, resulting in an amino-terminal cleavage product (nsp1) that contains both PCP1 domains. In contrast, due to the activity of PCP1α and PCP1β, the equivalent LDV and PRRSV proteins are cleaved into nsp1α and nsp1β. In both PCP1α and PCP1β, but not in CP2, the catalytic Cys (Fig. 6B, C) is immediately followed by a conserved aromatic residue. This sequence signature is a hallmark of cellular papain-like proteinases and was initially used to characterize viral proteinases as being 'papain-like'. Therefore, in the paper describing the identification of CP2 (Snijder *et al.*, 1995), the enzyme was distinguished from PCP1α and PCP1β. Subsequently, however, a distant variant of the papain-like fold was identified in a human ubiquitin carboxyl-terminal hydrolase (Johnston *et al.*, 1997) in which, as in the arterivirus CP2 (Fig. 6C), the conserved aromatic residue is replaced by Gly. Hence, the original reasons for discriminating between CP2 and PCP1α/PCP1β no longer exist. CP2 may adopt a papain-like fold and, if this is confirmed, its name should be modified accordingly.

Each of the three groups of arterivirus cysteine proteinases has a characteristic set of conserved residues that are scarcely overlapping (Fig. 6B, C). For example, the comparison of the most closely related pair of arterivirus cysteine proteinase groups (PCP1α/PCP1β) revealed that only one residue, the catalytic His, is absolutely conserved (Fig. 6B, C). (This number increases to four residues if the defective EAV PCP1α is not taken into account). All arterivirus accessory proteinases, with the exception of the inactivated EAV PCP1α, process (or are predicted to process) a single, downstream cleavage site (Fig. 2).

The coronaviruses MHV, HCoV and TGEV encode two accessory proteinases (Fig. 2). They are called coronavirus papain-like proteinases 1 and 2, PL1^{pro} and PL2^{pro}, respectively. IBV encodes only one accessory proteinase. It is called coronavirus papain-like proteinase, PL^{pro} (Fig. 2). Despite the relationship implied in the names, only a marginal similarity between the coronaviral and prototypic cellular proteinases is evident in an alignment based on the comparison of (predicted) secondary structures (Herold *et al.*, 1999; Fig. 6A). However, a statistically reliable, albeit local, primary structure similarity was detected between coronavirus PL^{pro} domains and the leader proteinase (Lpro) of foot-and-mouth disease virus (FMDV), a picornavirus (Gorbalenya *et al.*, 1991; A. E. Gorbalenya, unpublished data). This similarity was used to identify a papain-like catalytic centre in the FMDV Lpro, which was subsequently confirmed by mutagenesis experiments (reviewed in Ryan & Flint, 1997) and X-ray crystallography (Guarné *et al.*, 1998). Obviously, this relationship suggests that FMDV Lpro and coronavirus PL^{pro} domains have evolved under a similar selective pressure that separates them from other RNA virus papain-like proteinases including those of arteriviruses (for a list of these proteinases see Gorbalenya & Snijder, 1996). It is conceivable that common phenotypic features exist that are conserved in Lpro and PL^{pro} domains. It is also worth noting that the functional separation of coronavirus PL^{pro} domains and 3CL^{pro} into accessory and main proteinases also applies to the FMDV Lpro and 3Cpro, respectively. Thus, it is justified to treat the FMDV Lpro as the prototypic *viral* proteinase for coronavirus PL^{pro} domains (Table 1).

Comparative sequence analyses of the coronavirus accessory proteinases do not provide definitive support for the clustering of the PL1^{pro} and PL2^{pro} domains into two groups (as their names appear to imply) or for the association of the IBV PL^{pro} with one of these groups (A. E. Gorbalenya, unpublished observations). Indeed, the alignment of PL1^{pro} and PL2^{pro} (Fig. 6A) shows that only very few conserved residues are present exclusively in one of the two groups. Instead, and despite the low overall level of similarity in pairwise comparisons (13–32% identical residues), coronavirus accessory proteinases, including the IBV PL^{pro}, have eight absolutely conserved residues (Herold *et al.*, 1999). Among these are the catalytic Cys-His dyad as well as three Cys residues that are involved in the formation of a zinc-binding finger. The activity or activities of this zinc finger must be both essential and compatible with the different functions commonly found in related, but paralogous proteinases (i.e. enzymes that evolved by duplication rather than speciation). The conservation of this structural element embedded in the central region of these enzymes clearly discriminates the accessory proteinases of coronaviruses from their arterivirus counterparts.

The presence of only one PL^{pro} domain in IBV is most intriguing and can be interpreted in different ways. For example, in the course of coronavirus evolution, a duplication of the PL^{pro} domain might have occurred after the divergence of IBV from the rest of the coronavirus family. This scenario would imply that the PL^{pro} duplications have occurred independently in the arterivirus and coronavirus lineages. Alternatively, a (single) duplication could have taken place in a common nidovirus progenitor. In this case, the second PL^{pro} domain in IBV must have been deleted or diverged beyond recognition. Thus, the IBV PL^{pro} would be orthologous with either the PL1^{pro} or the PL2^{pro} group (i.e. these enzymes would have diverged from a common ancestor by speciation rather than by duplication). It has been noted before (Gorbalenya *et al.*, 1991) that the IBV PL^{pro} and the PL2^{pro} domains of the other coronaviruses are collinear in the replicase polyproteins (Fig. 2A), which favours the hypothesis that they might be orthologous proteins.

Based on a limited sequence similarity with a streptococcal cysteine proteinase, which belongs to a prokaryotic subset of

papain-like proteinases, the existence of another accessory proteinase in IBV was initially postulated (named *Streptococcus pneumoniae*-like proteinase, SPL) (Gorbalenya *et al*., 1989*b*). SPL was predicted in a region of pp1a/pp1ab that partly overlaps with the PL^{pro} domain identified 2 years later (Gorbalenya *et al*., 1991; Lee *et al*., 1991). The predicted catalytic Cys and His residues of SPL are not conserved in other coronavirus replicase polyproteins and, furthermore, the relevant Cys residue immediately follows the catalytic His residue of PL^{pro}. To our knowledge, such an overlapping organization of the active sites of two different enzymes has not been observed elsewhere and, clearly, this raises doubts about the correct identification of SPL and PL^{pro}. Given the conservation of PL^{pro} in all coronaviruses and the solid experimental support for its existence, it is safe to assume that the SPL domain is not functional.

Processing of the amino-proximal region of the coronavirus replicase polyproteins by accessory papain-like cysteine proteinases

The first data on the processing of coronavirus replicase polyproteins were obtained by translation of genomic MHV RNA in rabbit reticulocyte lysates. In pulse–chase experiments, a 28 kDa protein, p28, which is initially synthesized as part of a larger precursor protein, was identified (Denison & Perlman, 1986). This protein was also detected in virus-infected cells (Denison & Perlman, 1987). The genesis of p28 has been analysed in detail. First, *in vitro* translation experiments, combined with peptide mapping, were used to show that p28 represents the amino-terminal polypeptide (Soe *et al*., 1987). Second, it was shown that the proteolytic activity involved is virus-encoded and maps to the predicted $PL1^{pro}$ domain (Baker *et al*., 1989, 1993; Gorbalenya *et al*., 1991). Third, the scissile bond that is cleaved to release the carboxyl terminus of p28 was found to be located between Gly-247 and Val-248 (Dong & Baker, 1994; Hughes *et al*., 1995).

In subsequent studies, immunoprecipitation experiments with region-specific antisera revealed that additional proteolytic cleavages within the amino-proximal region of the MHV-A59 replicase polyproteins generate polypeptides of 65, 50, 240 and 290 kDa (p65, p50, p240 and p290) (Denison *et al*., 1992, 1995). Pulse–chase experiments indicated that the 290 kDa protein is the precursor of the 50 kDa and 240 kDa proteins (Denison *et al*., 1992). The experimental data also suggested that p65 is adjacent to the carboxyl terminus of p28 (Denison *et al*., 1995). Subsequently, a second $PL1^{pro}$ cleavage site was identified (Bonilla *et al*., 1995) and characterized by amino acid sequence analysis (Bonilla *et al*., 1997). This cleavage occurs at the Ala-832|Gly-833 peptide bond and it was concluded that p65 encompasses the MHV-A59 ORF1a-encoded amino acids 248–832.

A slightly different processing pattern has been reported for the JHM strain of MHV (Baker *et al*., 1989; Gao *et al*., 1996; Schiller *et al*., 1998). In this case, cleavage products of 28, 72 and 250 kDa were identified. Pulse–chase experiments revealed that the 72 kDa protein, p72, is further processed to a 65 kDa protein, p65 (Gao *et al*., 1996), whereas the 250 kDa protein, p250, is processed to a 210 kDa protein, p210 (Schiller *et al*., 1998). In the same study, another p250-derived, 40 kDa processing product, p40, was described, but a clear precursor–product relationship between p250 and p40 has not yet been established. Furthermore, the JHM-specific polypeptides, p250, p210 and p65, were found to comigrate with the A59-specific polypeptides, p290, p240 and p65, and, therefore, it is reasonable to assume that they represent identical proteins. Schiller *et al*. (1998) also concluded from these data that the precursor of p65 is p72 and it is unique to MHV-JHM. They also propose that the carboxyl-terminal cleavage site of p72 is the peptide bond between Gly-904 and Val-905.

Information on the proteolytic processing of the amino-proximal region of the replicase polyproteins of HCoV and IBV is quite limited. For both viruses, only the amino-terminal cleavage products have been characterized so far. By using a monospecific antiserum directed against the HCoV ORF1a-encoded amino acids 41–250, a polypeptide with an apparent molecular mass of 9 kDa, p9, was identified in HCoV-infected cells (Herold *et al*., 1998). This protein, p9, was shown to be generated by $PL1^{pro}$-mediated cleavage between Gly-111 and Asn-112, which is equivalent to the p28 cleavage site recognized by the MHV $PL1^{pro}$. In the same study, polypeptides with apparent molecular masses of 93 kDa (p93), 170 kDa (p170) and 230 kDa (p230) were specifically immunoprecipitated from lysates of metabolically labelled, HCoV-infected cells. The identity of these proteins and potential precursor–product relationships remain to be determined.

For IBV, the amino-terminal processing product of the replicase polyproteins has been suggested to be a protein with an apparent molecular mass of 87 kDa (Lim & Liu, 1998). Conflicting data have been published with respect to the proteinase responsible for the processing of the 87 kDa protein. Thus, in an initial report, the involvement of a cellular proteinase in the production of p87 *in vitro* was proposed (Liu *et al*., 1995). Recently, however, it was shown that mutagenesis of the putative active-site residues of the previously predicted PL^{pro} domain (Gorbalenya *et al*., 1991; Lee *et al*., 1991) completely abolished the production of the 87 kDa protein in a Vero cell transient expression system (Lim & Liu, 1998). This result and additional data presented in this paper suggest that the production of p87 is mediated by the virus-encoded PL^{pro}, rather than a cellular proteinase. In the same study, site-directed mutagenesis of a candidate PL^{pro} cleavage site suggested that p87 is released by cleavage of the Gly-673|Gly-674 peptide bond. Interestingly, no evidence has been obtained for the existence of an additional, upstream cleavage site that would be equivalent to the sites involved in the release of p28 and p9 from the MHV and HCoV replicase polyproteins, respectively.

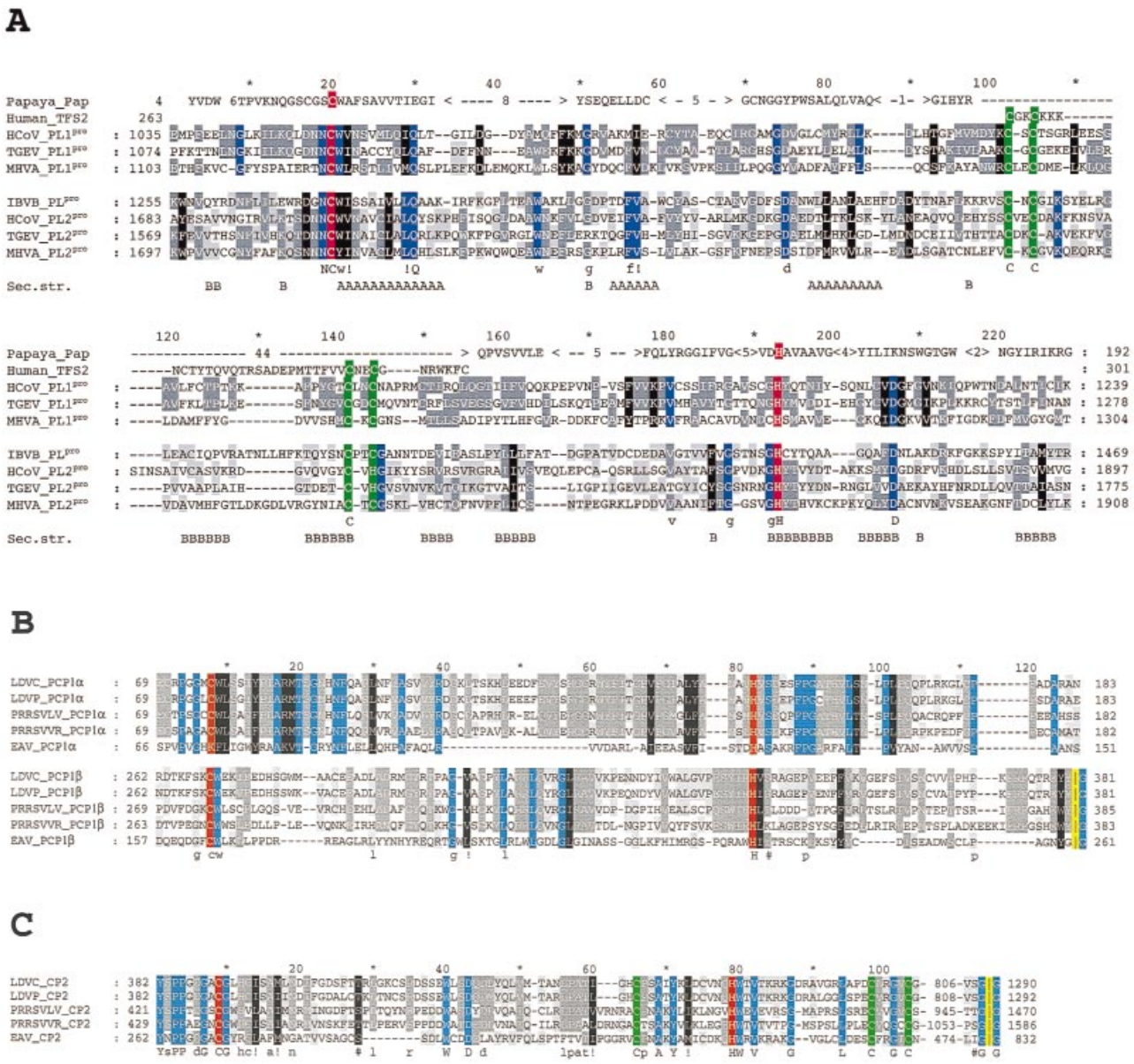

Fig. 6. Multiple alignments of nidovirus accessory proteinases. The alignment of the coronavirus papain-like proteinase domains, PL^{pro}, $PL1^{pro}$ and $PL2^{pro}$ (A), together with papain and human transcription factor S2 (TFS2), was produced as described in Herold *et al.* (1999). The alignment also features secondary structure elements of papain and TFS2 (A, α-helix; B, β-strand) which may also be conserved in coronavirus PL^{pro} domains. The multiple alignments of arterivirus PCP1α and PCP1β (B) and CP2 (C) were produced using the ClustalX program (Thompson *et al.*, 1997). The green background in several positions of the coronavirus papain-like proteinase (A) and CP2 (C) alignments highlights residues thought to bind Zn^{2+} (Snijder *et al.*, 1995; Herold *et al.*, 1999). For additional details see legend to Fig. 3.

Enzymatic and structural properties of coronavirus accessory proteinases

On the basis of previous predictions (Gorbalenya *et al.*, 1991; Lee *et al.*, 1991; Herold *et al.*, 1993), mutagenesis studies of putative Cys-His catalytic dyads of three different coronavirus PL^{pro} domains were done. The available data suggest that Cys-1137/His-1288, Cys-1054/His-1205 and Cys-1274/His-1437 are the putative catalytic residues of the MHV-JHM $PL1^{pro}$, the HCoV $PL1^{pro}$ and the IBV PL^{pro}, respectively (Baker *et al.*, 1993; Herold *et al.*, 1998; Lim & Liu, 1998). The Cys-1137/His-1288 dyad of MHV-JHM corresponds to Cys-1121/His-1272 in the MHV-A59 replicase polyproteins (Bonilla *et al.*, 1994) and the involvement of His-1272 in proteolysis has been experimentally confirmed (Bonilla *et al.*, 1995). Recently, the papain-like fold of the coronavirus

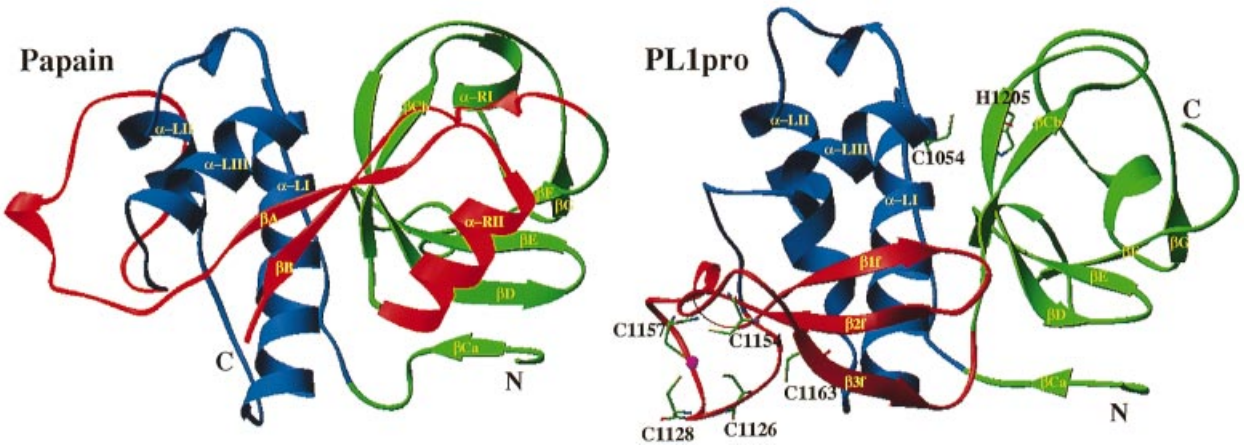

Fig. 7. A crude structural model of HCoV PL1^{pro}. The PL1^{pro} model was generated using two structural templates, papain (pdb ID: 1ppn) and TFS2 (1tfi), and followed the alignment in Fig. 6 (A) as described in Herold *et al.* (1999). The structure of papain is shown for comparison. The structures of both proteinases are displayed as ribbons (Carson, 1997) and are split into three domains. These are coloured as follows: blue, left-hand α-helix domain; green, right-hand β-sheet domain without counterparts of βA- and βB-strands; red, zinc finger domain in PL1^{pro} and the inter-domain loop along with βA- and βB-strands and α-RII-helix in papain. Cysteine residues of the zinc finger, including the non-essential C1163, as well as the catalytic dyad residues of PL1^{pro} are shown in the ball-and-stick representation. Zn^{2+} is also shown. The picture is reproduced from Herold *et al.* (1999) with permission.

accessory proteinases was modelled in some detail (Herold *et al.*, 1999). In this analysis, a number of additional residues were postulated to be involved in the formation of the active site of coronavirus PL^{pro} domains. These residues were probed by point mutations in the HCoV PL1^{pro} and shown to be essential for proteolytic activity. However, their precise roles remain to be elucidated.

A theoretical analysis of the primary and (predicted) secondary structures of coronavirus papain-like proteinase domains has also recently delineated a conserved zinc finger structure. This structure is organized as a separate, (possibly) β-sheet domain, connecting the left- and right-hand domains of a papain-like fold in an unprecedented way (Figs 6A and 7) (Herold *et al.*, 1999). Several striking implications of this analysis have been tested experimentally using the HCoV PL1^{pro}, expressed and purified as an *E. coli* maltose-binding fusion protein (MBP–PL1^{pro}). First, using a spectrometric method, equimolar binding of Zn^{2+} by MBP–PL1^{pro} was documented. Second, denaturation/renaturation experiments using MBP–PL1^{pro} revealed a strong Zn^{2+}-dependence for proteolytic activity and a tight association of Zn^{2+} with PL1^{pro}. And third, the replacement of the four cysteine residues predicted to coordinate Zn^{2+} resulted in a selective inactivation of the enzyme as judged by *in vitro cis-* and *trans*-cleavage assays. Thus, it was concluded that the Zn^{2+} ion is tetrahedrally coordinated by Cys-1126, Cys-1128, Cys-1154 and Cys-1157 of the HCoV PL1^{pro}. The zinc finger domain replaces a poorly conserved structure of variable size which connects the two domains of the papain fold in numerous papain-like proteinases (Fig. 7). For example, the analogous structure in the FMDV Lpro is composed of only two parallel β-strands (Guarné *et al.*, 1998). This accounts, at least partially, for the smaller size of this picornavirus proteinase compared with the coronavirus homologues.

The available data are also consistent with an essential, structural, rather than catalytic, role for the zinc finger. However, it is not clear how this finger actually modulates the proteolytic activity of PL1^{pro}. The results of the mutagenesis study and the current low-resolution model of PL1^{pro} (Fig. 7) would allow for several mechanisms, including control of the interdomain motions essential for catalysis or direct interactions between the zinc finger and the substrate or the structure(s) forming the PL1^{pro} active site. Furthermore, Herold *et al.* (1999) speculate that the predicted, prominent position of the zinc finger, in front of the substrate-binding cleft, make it an excellent candidate for coupling external signals with the regulation of PL1^{pro}-mediated processing. Keeping in mind that the zinc finger domain was modelled to adopt the zinc-ribbon architecture conserved in a number of cellular transcription factors, it is reasonable to think that these external signals might, for example, originate from (viral) RNA or the proteins of replication/transcription complexes (Herold *et al.*, 1999).

Sequence analyses of three cleavage sites have provided information on the substrate specificity of PL1^{pro}. For both of the MHV PL1^{pro} cleavage sites identified so far, Gly-247|Val-248 and Ala-832|Gly-833, extensive mutation studies have been done (Dong & Baker, 1994; Hughes *et al.*, 1995; Bonilla *et al.*, 1997). A number of conclusions can be drawn. First, the most important structural element of MHV PL1^{pro} cleavage sites appears to be the P1 residue. This position is invariably occupied by Gly or Ala, with Gly being

the clearly preferred residue. Second, only an extremely limited number of amino acid substitutions are tolerated at the P2 position. For example, in the case of the Gly-247|Val-248 cleavage site, nine different substitutions of P2 obviated the function of the cleavage site (Dong & Baker, 1994; Hughes *et al.*, 1995). Third, although the P1′ position tolerated a quite large number of substitutions, a clear preference for small, uncharged amino acids was found. Again, Gly and Ala were the most active amino acids in different assays. Fourth, a basic residue is, most likely, part of the substrate signature because substrates with either Arg or Lys at the P5 position and Arg at the P2 position gave high cleavage rates. In this respect, it is also noteworthy that the presence of Arg at the P2 position was sufficient to compensate for an inactivating Arg-by-Met substitution at the P5 position (Bonilla *et al.*, 1997). Finally, the data indicate that the positions P3, P4, P2′ and P3′ tolerate a large number of amino acid substitutions. Therefore, it is thought that these residues do not contribute significantly to substrate recognition.

The substrate specificity of the HCoV PL1^{pro} was determined by sequence analysis of the carboxyl-terminal cleavage site of p9 (Herold *et al.*, 1998). By and large, the cleavage site corresponds well to the general pattern outlined above for the MHV PL1^{pro}. Again, a Gly residue was found at the P1 position and the P1′ and P5 positions were occupied by small uncharged (Asn-112) and basic (Lys-107) residues, respectively. Furthermore, sequence alignment revealed that, within the replicase polyproteins of three different coronaviruses, the relevant PL1^{pro} cleavage site is an integral part of the only significantly conserved sequence block upstream of PL1^{pro}. An analysis of this block suggested that the PL1^{pro} cleavage site might have migrated by two residues in either the MHV or HCoV/TGEV evolutionary lineage from its initial position in the common coronavirus ancestor. Such an event would have taken place under a complex and evolving selective pressure upon this locus. For the recently described IBV PL^{pro} cleavage site (Lim & Liu, 1998), site-directed mutagenesis suggests an involvement of the pp1a/pp1ab amino acids Gly-673 and Gly-674 in the release of the p87 carboxyl terminus. However, in the absence of protein sequence data it is still premature to draw conclusions about the IBV PL^{pro} substrate specificity.

The first detailed reports of PL1^{pro} activity detected no *trans* activity for the MHV-JHM enzyme using an *in vitro* translation assay (Baker *et al.*, 1989). Similarly, the IBV PL^{pro} was not active *in trans* upon transient expression in Vero cells (Lim & Liu, 1998). Recently, however, evidence has been presented that two other related PL1^{pro} domains, those of MHV-A59 and HCoV, are active *in trans*, at least *in vitro* (Bonilla *et al.*, 1997; Herold *et al.*, 1998). These and other results (Baker *et al.*, 1993; Bonilla *et al.*, 1995; Teng *et al.*, 1999) suggest that a number of factors can profoundly modulate proteolytic processing by the PL1^{pro}, both *in cis* and *in trans*.

In MHV-A59, a domain‖ of 233 amino acids (pp1a/pp1ab residues 1084–1316) was shown to be required for the *cis* cleavage that generates p28 (Bonilla *et al.*, 1995). Subsequently, however, it was found that this form of the enzyme had a lower activity compared to a form containing amino acids 1062–1364. In addition, cleavage at both MHV-A59 PL1^{pro} processing sites seems to be modulated by sequences downstream of PL1^{pro}. Thus, for example, it was found that a polypeptide encompassing amino acids Lys-869 to Pro-2028 showed a more than fivefold enhanced cleavage activity compared to a polypeptide encompassing Lys-869 to Gln-1314 (Teng *et al.*, 1999). The region absent in the smaller protein contained the X domain, which is highly conserved among *Togaviridae* and *Coronaviridae* (Gorbalenya *et al.*, 1991), and the putative PL2^{pro} domain (Lee *et al.*, 1991; Bonilla *et al.*, 1994) (Fig. 2). Thus, it is tempting to speculate that sequences downstream of PL1^{pro} might be involved in the regulation of its activity, as has been previously suggested (Gorbalenya *et al.*, 1991). The question of whether these domains provide specific activities or 'simply' affect the overall conformation of the protein has yet to be answered. Intriguingly, no evidence for a proteolytic activity of the PL2^{pro} domain has been obtained so far (Gao *et al.*, 1996; Teng *et al.*, 1999).

With respect to PL1^{pro} substrate(s), it has been shown that the production of p28 requires the presence of downstream sequences (Baker *et al.*, 1993; Bonilla *et al.*, 1995). This observation was confirmed and extended by a study in which the purified PL1^{pro} domain (amino acids 1062–1364) was assayed *in trans* using a set of *in vitro*-translated, ORF1a-derived substrates (Teng *et al.*, 1999). The data showed that a minimum of 622 amino acids from the amino-proximal region of the replicase polyproteins are required to obtain high rates of cleavage by PL1^{pro}. Consistently, recombinant forms of PL1^{pro} failed to cleave synthetic peptides containing cognate cleavage sites (Teng *et al.*, 1999). It can be concluded that the region delimited by amino acids 302–622 (or even amino acids further downstream) specifically contributes to proteolysis at the p28 site, probably by enhancing presentation of the substrate or activating the enzyme.

Only very limited information is currently available about the requirements of the accessory proteinase-mediated cleavages in other coronaviruses. In HCoV, as in MHV, the sequence between the p9 cleavage site (which is equivalent to the MHV p28 cleavage site) and the catalytic domain was shown to contain determinant(s) required for PL1^{pro}-mediated processing *in trans* in reticulocyte lysates (Herold *et al.*, 1998). In IBV, the situation appears to be different. It was shown, in a recombinant vaccinia virus/T7 expression system, that cleavage at the p87 site *in cis* tolerates a large deletion of

‖ This domain and the corresponding domain of HoCV (Herold *et al.*, 1998) could be called 'minimal'. We avoid using this word as it may give the impression that the domain in question is necessary and sufficient for proteolytic activity. However, as discussed, the presence of additional domains in the proteinase assay needs to be considered in the interpretation of the results.

approximately half the region between the cleavage site and the catalytic domain of PL^{pro}. Also, again in contrast to the MHV data, the proteolytic reaction of PL^{pro} appeared to be insensitive to the presence of approximately 300 amino acids flanking the catalytic domain on its carboxyl terminus (Lim & Liu, 1998). The above data suggest that the dependence of PL^{pro}-mediated cleavages upon specific sequences, which flank the cleavage site(s) and the catalytic domain, may be conserved in some, but not all, coronaviruses.

In common with the coronavirus main proteinases, the coronavirus accessory proteinases have also been shown to deviate significantly from the prototypic enzymes. The availability of purified, recombinant $PL1^{pro}$ domains from two coronaviruses (Herold *et al.*, 1999; Teng *et al.*, 1999) should now make these enzymes amenable to detailed structural analyses. This information should help in understanding the complex organization of viral proteins carrying PL^{pro} domains and provide insights to their structure–function relationships.

Arterivirus nsp1 papain-like proteinases

The activity of *in vitro*-translated EAV nsp1 papain-like cysteine proteinase (PCP1*β*, the origin of the suffix, *β*, is clarified below) facilitated its early experimental characterization. Comparative sequence analysis had predicted the presence of a cysteine proteinase, with limited sequence similarity to cellular and other viral papain-like cysteine proteinases, in the amino-proximal region of the EAV replicase polyproteins (den Boon *et al.*, 1991). Upon translation of RNA transcripts encoding this region, a 29 kDa, amino-terminal cleavage product was rapidly generated (Snijder *et al.*, 1992). Uncleaved precursor proteins could not be detected and pulse–chase studies suggested the almost immediate (co-translational) release of an amino-terminal product, which was named nsp1 (Fig. 2B). The same protein was also detected in virus-infected cells, where it appeared to be produced with similar kinetics (Snijder *et al.*, 1994). Furthermore, the EAV nsp1 proteinase was found to be active in *E. coli* when expressed as part of a bacterial fusion protein (Snijder *et al.*, 1992). Amino-terminal sequence analysis of an *E. coli*-derived cleavage product showed that the nsp1|2 scissile bond was located between Gly-260 and Gly-261 of the EAV ORF1a and ORF1ab polyproteins.

The subsequent sequence analyses of the genomes of two other arteriviruses, LDV (Godeny *et al.*, 1993; Palmer *et al.*, 1995) and PRRSV (Lelystad strain; Meulenberg *et al.*, 1993) revealed that their nsp1 region was substantially larger than that of EAV (about 380 versus 260 residues). Furthermore, this region was found to contain two, instead of one, PCP domains. The sequence analysis also revealed an overall collinearity between the amino-terminal regions of EAV and PRRSV/LDV (den Boon *et al.*, 1995). Specifically, the carboxyl-terminal proteinase domain of PRRSV/LDV could be aligned with the EAV nsp1 PCP1*β*, whereas the amino-terminal domain of PRRSV/LDV matched with a PCP remnant that was identified upstream of the EAV PCP1*β* in nsp1. This new domain was named PCP1*α*. The functionality of both of these domains was experimentally confirmed for PRRSV/LDV by den Boon *et al.* (1995). First, PCP1*α* was shown to mediate the rapid liberation of an amino-terminal, 20–22 kDa cleavage product (nsp1*α*) (Fig. 2B). Second, the PRRSV/LDV PCP1*β* was shown to act like its EAV orthologue in that it cleaves the nsp1|2 junction. Thus, both PCP1*α* and PCP1*β* contribute to the production of the nsp1*β* protein (Fig. 2B) (den Boon *et al.*, 1995).

The putative PCP1*α* active-site residues were shown to be Cys-76/His-146 and Cys-76/His-147 in PRRSV and LDV, respectively, but a conserved putative nsp1*α*|1*β* cleavage site could not be identified (Godeny *et al.*, 1993; Meulenberg *et al.*, 1993; den Boon *et al.*, 1995). Although a Lys residue was found in place of the presumed catalytic Cys residue, which explains the proteolytic deficiency of the EAV PCP1*α* domain, the sequences around EAV His-122 display convincing sequence similarity with the region surrounding the PCP1*α* active-site His-146/147 of LDV/PRRSV (Fig. 6B). The consequences of the inactivation of the EAV PCP1*α* domain are unclear but the partial conservation of the inactivated PCP1*α* suggests that this part of the replicase probably contains an additional, non-proteolytic function that is conserved in all arteriviruses (den Boon *et al.*, 1995; Gorbalenya & Snijder, 1996).

The identification of Cys-164 and His-230 as the probable EAV PCP1*β* catalytic dyad (den Boon *et al.*, 1991) was supported by data from site-directed mutagenesis (Snijder *et al.*, 1992) which showed that the replacement of the putative active-site residues Cys-164 (by Ser or Gly) or His-230 (by Val, Ala or Gly) completely inactivated proteinase function. Deletion mutagenesis was used to delimit the minimal domain required for activity to residues 123–263. The putative PCP1*β* active-site residues in PRRSV and LDV are Cys-276/His-345 and Cys-269/His-340, respectively (den Boon *et al.*, 1995). Sequence comparison suggests that the PCP1*β* domains of both PRRSV and LDV cleave between Tyr and Gly (Tyr-384|Gly-385 and Tyr-380|Gly-381, respectively; Fig. 6B).

Attempts to achieve cleavage *in trans* using arterivirus PCPs were unsuccessful (den Boon *et al.*, 1995; Snijder *et al.*, 1992). A similar situation was also observed for another virus proteinase, the Sindbis virus capsid protein, which was shown to act exclusively *in cis* (Choi *et al.*, 1991). In the case of Sindbis virus, the carboxyl terminus of the proteinase was found to remain in the P1 substrate site subsequent to the autocatalytic *cis* cleavage of the capsid protein. The structure analysis of this protein revealed that the carboxyl-terminal Trp residue seals the active site of the proteinase, rendering the enzyme inactive after the release of the downstream protein. An analogous organization may be adopted by the arterivirus PCPs, which would explain their readily detectable *cis* activity and the absence of *trans* activity. A limited mutagenesis study of the nsp1|2 cleavage site revealed that its P1 position is more

sensitive to replacements than its P1′ position and that sequences downstream of P2′ are not required for processing (Snijder *et al.*, 1992).

Arterivirus nsp2 cysteine proteinase

The arterivirus CP2 is the most carboxyl-terminally located member of the array of three cysteine proteinase domains present in the amino-terminal 500 residues of the replicase polyproteins (Fig. 2B). CP2 is located in the amino-terminal region of nsp2 and is highly conserved among arteriviruses, although it has only been studied experimentally for EAV (Snijder *et al.*, 1995). The cleavage of the nsp2|3 junction appears to be the single processing step mediated by CP2. The size of the resulting nsp2 cleavage product is quite variable, ranging from 571 residues in EAV to 1195 residues in the VR2332 strain of PRRSV (Nelsen *et al.*, 1999). For EAV, it has been shown that cleaved nsp2 is an essential co-factor for cleavage of the nsp4|5 site by the nsp4 3CLSP (Wassenaar *et al.*, 1997).

The activity of the EAV nsp2 CP2 has been analysed *in vivo* using infected cells and eukaryotic expression systems (Snijder *et al.*, 1994, 1995), where the proteinase cleaves rapidly and probably *in cis*. *Trans*-cleavage activity could also be demonstrated in a eukaryotic expression system, albeit with relatively low efficiency (Snijder *et al.*, 1995). Because its activity could not be demonstrated *in vitro* there is only indirect evidence to support the idea that the EAV CP2 cleaves between two Gly residues (Gly-831|Gly-832). Specifically, it has been shown that mutagenesis of the putative P1 residue (Gly-831 to Pro) abolished processing of the nsp2|3 site (Snijder *et al.*, 1996). Also, the proposed site is conserved among all arteriviruses (Fig. 6C).

Comparative sequence analysis and site-directed mutagenesis have characterized the CP2 as a cysteine endopeptidase whose conserved domain encompasses about 100 residues. Like the PCP domains, this viral enzyme most clearly resembles viral papain-like cysteine proteinases. Residues Cys-270 and His-332 are assumed to form the CP2 catalytic dyad because their replacement completely inactivated proteolytic activity (Snijder *et al.*, 1995). Although the distance between these active-site residues resembles that found in arterivirus papain-like proteinases (Fig. 6B, C), there is one important difference. The putative catalytic Cys-270 is flanked by Gly-271, and not by a bulky, hydrophobic residue which is almost always found at this position in viral proteinases and is a hallmark of this group of proteinases (Gorbalenya & Snijder, 1996; although see above and Johnston *et al.*, 1997). An attempt to convert CP2 into an active 'canonical' papain-like proteinase through the replacement of Gly-271 by Trp failed (Snijder *et al.*, 1995), indicating that the fixation of Gly in CP2 is likely to be tightly coupled with additional changes. Unlike the two other arterivirus papain-like proteinases and the FMDV Lpro, CP2 does not cleave immediately downstream of the active-site His. Indeed, about 500 residues separate CP2 and its cleavage site in EAV and this distance can even extend to more than 1100 residues in PRRSV-VR2332 (Nelsen *et al.*, 1999) (Fig. 6C). The entire CP2 domain is highly conserved among arteriviruses, which is a remarkable difference to the PCP1α and PCP1β domains. Among the conserved residues are a number of cysteines and one aspartate residue (Fig. 6C). While the replacement of the conserved or neighbouring acidic residues did not influence CP2 activity, the substitution of the conserved Cys residues in EAV abolished (Cys-319, Cys-349 and Cys-354) or reduced (Cys-344 and Cys-356) processing at the nsp2|3 site (Snijder *et al.*, 1995). It is conceivable that the three conserved and essential Cys residues may be part of a zinc finger, which would resemble the situation found in coronavirus PL^{pro} domains where the structural importance of the zinc finger for the proteinase activity has recently been shown (Herold *et al.*, 1999; see also above).

The arterivirus nsp2 appears to be a multi-domain protein. It contains highly conserved regions, e.g. a domain with a number of conserved Cys residues in its carboxyl-terminal half, but there are also sequences that cannot be aligned among arteriviruses. A detailed description of the functions associated with these domains remains to be obtained, but it has recently become clear that nsp2 is involved in the generation of a membrane-associated replication complex (van der Meer *et al.*, 1998; Pedersen *et al.*, 1999). This may also have consequences for the proteolytic activity of CP2. Both large and small deletions in the EAV nsp2 region that separates the CP domain and its cleavage site were found to interfere with proteolytic processing at the nsp2|3 junction (Snijder *et al.*, 1995).

Proteinases as regulators of nidovirus replication

As discussed above, the accessory nidovirus proteinases mediate only very few cleavages in the relatively divergent, amino-terminal portion of the replicase polyproteins. In contrast, the main proteinases are responsible for the extensive proteolytic processing of the so-called 'core replicase' (Snijder & Spaan, 1995). The latter region encompasses all of the major conserved domains starting from the hydrophobic domain upstream of the main proteinase and extending to the carboxyl terminus of the replicase polyprotein (Fig. 2). Therefore, nidovirus proteinases are believed to play an important regulatory role in the generation (or inactivation) of specific protein functions at certain stages of the virus life-cycle. This controlled proteolysis is thought to be mainly determined by the substrate specificity of the proteinases and the accessibility of cleavage sites in the context of specific intermediate products. Indeed, peptides mimicking different cleavage sites recognized by the HCoV 3CL^{pro} were processed with remarkably different kinetics (Ziebuhr & Siddell, 1999). This observation is also in agreement with the information available

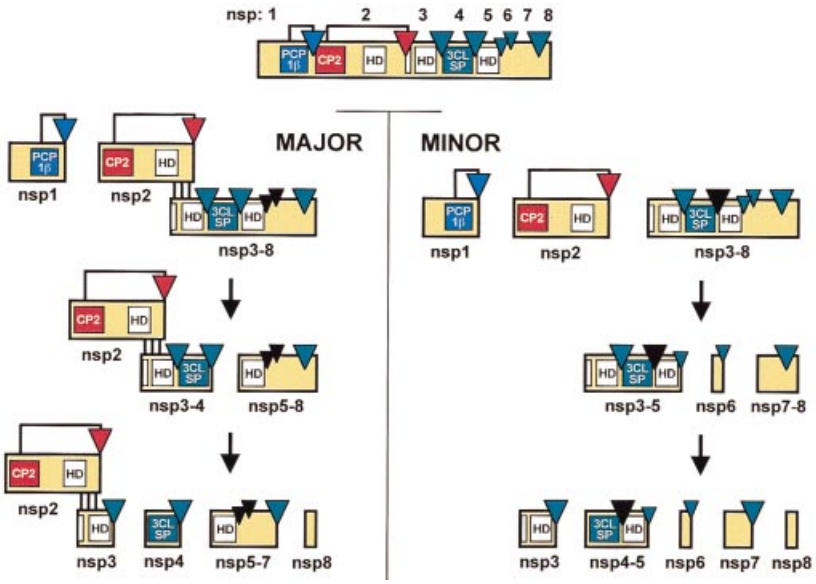

Fig. 8. Outline of the alternative proteolytic processing pathways for the replicase ORF1a polyprotein of the arterivirus EAV. The three EAV proteinases (PCP1β, CP2 and 3CLSP), the cleavage site map of the ORF1a protein and the EAV nsp nomenclature are shown at the top. Prominent hydrophobic domains (HD) in nsp2, nsp3 and nsp5 are indicated. The lower part of the figure shows the two alternative processing pathways relevant to the carboxyl-terminal half of the EAV ORF1a protein. Not all cleavage products are depicted. The association of cleaved nsp2 with nsp3–8 (and probably also nsp3–12) was shown to direct cleavage of the nsp4|5 site by the nsp4 3CLSP (major pathway). Alternatively, in the absence of nsp2, the nsp5|6 and 6|7 sites were found to be processed and the nsp4|5 junction remained uncleaved. The status of the small nsp6 subunit (fully cleaved or partially associated with nsp5 or nsp7) remains to be elucidated.

for EAV, which is currently the most advanced experimental system for the analysis of the proteinase-controlled regulation of the nidovirus life-cycle *in vivo*. Using reverse genetics, van Dinten *et al.* (1997, 1999) have shown for EAV that abolishment of the processing of the ORF1b-encoded portion of the ORF1ab polyprotein (by site-directed mutagenesis of cleavage sites) either severely affected or completely inhibited arterivirus replication. Preliminary results from a similar cleavage site mutagenesis study revealed the importance of proteolytic processing with regard to the ORF1a-encoded portions of the replicase polyproteins (M. A. Tijms & E. J. Snijder, unpublished results). Furthermore, rapid and slow cleavages have been identified in the EAV system (Snijder *et al.*, 1996). Thus, in pulse–chase experiments, cleavage of the nsp4|5 junction was rapid but many other cleavages were not even completed after a 6 h chase. In addition, two alternative processing pathways were elucidated for the carboxyl-terminal half of the ORF1a protein (Wassenaar *et al.*, 1997). As a result, many (alternative) processing intermediates are generated in infected cells, including a number of intermediates that contain 3CLSP itself. Whether these replicase-processing intermediates fulfil specific functions in the EAV life-cycle, as they do in certain other positive-stranded RNA viruses, is still an open question. However, this question can now be addressed using the recently developed infectious cDNA clones of EAV and PRRSV (van Dinten *et al.*, 1997; Meulenberg *et al.*, 1998).

As mentioned above, two alternative pathways can be followed for the processing of the carboxyl-terminal part of the EAV ORF1a protein (Fig. 8). Either the nsp4|5 ('major pathway') or the nsp5|6 and nsp6|7 ('minor pathway') sites are processed (Wassenaar *et al.*, 1997). Cleavage at either of these sites is believed to render the alternative site(s) non-accessible. It remains to be shown which factors determine the selection of the site that is initially cleaved by the 3CLSP. It is conceivable that the association of cleaved nsp2 with the nsp3–8 precursor triggers the 3CLSP to cleave the nsp4|5 site. In this complex, nsp2 is likely to have a strong interaction with nsp3 (Snijder *et al.*, 1994). If this model is correct, a decisive role in the major pathway could be assigned to CP2 (since it generates nsp2). Furthermore, it should be remembered that three processing end-products of the ORF1a polyprotein (nsp2, nsp3 and nsp5) contain hydrophobic domains that presumably mediate the anchoring of the EAV replication complex to intracellular membranes (van Dinten *et al.*, 1996; van der Meer *et al.*, 1998) and result in modification of these membranes into characteristic double-membrane vesicles (Pedersen *et al.*, 1999). Thus, it is reasonable to speculate that a specific, membrane-associated folding or post-translational organization of the nsp2–5 complex may determine the selection of a particular processing pathway.

So far, the functions of the two pathways in the EAV life-cycle remain completely obscure. However, some immediate consequences of the interplay between the two pathways are

evident. First, two proteins, nsp5 and nsp5–6, are apparently not produced at all since cleavage of the nsp4|5 site excludes cleavage of the nsp5|6 and nsp6|7 sites and vice versa. Theoretically, this strict down-regulation of two proteins could have evolved to prevent a deleterious effect of nsp5 or nsp5–6. This effect is reliably suppressed if these domains are expressed only as part of larger processing intermediates. Second, it is noteworthy that four domains, nsp4, nsp5, nsp6 and nsp7, reside in different end-products of the two pathways (Fig. 8). Two of these proteins (nsp5–7 of the major pathway and nsp4–5 of the minor pathway) contain the nsp5 hydrophobic moiety that is likely to mediate their membrane association. In other words, using two alternative processing pathways, the nsp5 domain may be capable of directing different domains (either nsp6–7 or nsp4) to the same cellular (membrane) compartment. The pathway-dependent targeting of these three domains may have a profound impact on the functions of the multi-domain complexes of which they are part. For instance, a permanent membrane association may selectively constrain the substrate availability for 3CLSP, which, when produced as the fully cleaved nsp4, was shown to be efficient in processing other parts of the replicase polyprotein *in trans* (van Dinten *et al.*, 1999).

The question of whether the two processing pathways of the carboxyl-terminal part of the ORF1a polyprotein are also used to process the corresponding part of the ORF1ab (frameshift) polyprotein remains to be addressed. The experiments that uncovered the two processing pathways were based on expression of only the EAV ORF1a polyprotein in the recombinant vaccinia virus/T7 expression system. This indicates that ORF1b-encoded sequences are not required for the utilization of alternative pathways. However, it cannot be excluded that the extension of the ORF1a protein with the ORF1b-encoded polypeptide might affect or modulate processing of this region.

Concluding remarks

For many years, the study of viral proteinases has produced a wealth of information on the structural constraints that are compatible with the function of a proteolytic enzyme. Although all these viral proteins necessarily meet the fundamental requirements for proteolytic activity, i.e. the presence of (i) a substrate-binding cleft, (ii) a reaction centre, and (iii) an appropriate mechanism for transition state stabilization, astonishing variations of the well-known structural patterns as well as unique, virus-specific folds have emerged (Babé & Craik, 1997). As discussed above, the nidovirus proteinases are part of this diversity and possess a number of properties that, in many respects, discriminate them from related proteinases of cellular and viral origin. It can be expected that the detailed analysis of nidovirus proteinases will substantially broaden our horizons concerning the protein structures that have evolved to optimize and extend the functions of proteolytic enzymes involved in virus replication.

The diversity of nidovirus proteinases is so pronounced that even the most sensitive computer-aided methods do not reveal any specific relationship between arterivirus and coronavirus enzymes that would justify their grouping and separation from other proteolytic enzymes. However, this dissimilarity at the sequence level is in sharp contrast to the collinearity of the arrays of conserved replicative domains (including the proteinases) and to the similar distribution of proteinase cleavage sites in the nidovirus replicase polyproteins. Thus, the grouping of the nidovirus proteinases is essentially based on the organization and expression of the replicase polyproteins and not on sequence similarity of the proteolytic enzymes. Despite the lack of evidence from structural studies, the hypothesis of divergent evolution from a common nidovirus ancestor, which contained a replicase (gene) with an organization resembling that of the contemporary nidoviruses, provides the most parsimonious explanation for the observed diversity of the proteolytic enzymes. It remains a goal for future studies to identify the selection pressures that, in the main nidovirus lineages, have driven the profound divergence of the proteinases and their targets, as well as the selective forces that determined the conservation of these specific patterns during evolution. As a step toward this goal, a comprehensive characterization of the proteolytic enzymes of the toroviruses, the third genus within the *Nidovirales*, should be conducted (for a review on toroviruses, see Snijder & Horzinek, 1993). With respect to genome size, toroviruses are intermediate between arteriviruses and coronaviruses, and, possibly, the torovirus proteinases bridge the huge gap that separates the proteolytic enzymes of the other two nidovirus genera. The sequence data currently available for toroviruses does not include proteinases, and, furthermore, no experimental data on the processing of the replicase polyproteins have been reported. However, based on the substrate specificity of nidovirus main proteinases and the alignment of the nidovirus replicase polyproteins (A. E. Gorbalenya, unpublished data), a number of potential cleavage sites can be discerned in the published sequence (Snijder *et al.*, 1990) of the ORF1b-encoded region of the replicase polyprotein of Berne virus.

As outlined in this review, the study of the processing pathways of the nidovirus replicase polyproteins led to the identification of a large number of intermediate and end-products that are believed to represent components of the viral replication complex. These results provide a solid basis for the elucidation of individual protein functions involved in various stages of the nidovirus life-cycle. Obviously, apart from the proteinases themselves, the initial studies were focussed on proteins harbouring putative replicative functions (Heusipp *et al.*, 1997*b*; van Dinten *et al.*, 1997, 1999) or functions related to the localization of the replication complex. Given the organization and extraordinary size of nidovirus replicase

polyproteins, it can be expected that the assembly and function of the viral replication machinery is considerably more complex than that of most other RNA viruses. For example, it has been speculated that *Nidovirales* and plant closteroviruses (the largest plant positive-stranded RNA viruses with genome sizes up to 20 kb) have evolved specific strategies to build and maintain large RNA genomes (Dolja *et al*., 1994; Agranovsky, 1996). For nidoviruses, only a limited number of protein domains encoded by the replicase gene have been correlated with specific activities or functions. The larger part of the replicase gene encodes proteins for which no counterparts have been identified in other cellular or viral systems. For example, the portion of the nidovirus replicase polyproteins between the main proteinase and the carboxyl terminus of pp1a is extensively processed to produce up to six proteins (Wassenaar *et al*., 1997; Liu *et al*., 1997; Lu *et al*., 1998; Ng & Liu, 1998; Ziebuhr & Siddell, 1999). For most of them, however, no functional assignments have been made. Without doubt, reverse-genetic systems, at the moment restricted to arteriviruses, will prove to be valuable tools in the investigation of specific functions involved in different aspects of virus replication and transcription.

Nidovirus-encoded proteinases, like many other viral proteinases, are highly effective regulators of virus replication and, indirectly, possibly even of virion biogenesis (van Dinten *et al*., 1999). Consequently, they represent ideal targets for therapeutic intervention. This would be highly desirable for a number of economically important nidovirus infections. The elaborated substrate specificity of the nidovirus main proteinases should facilitate the identification of selective low molecular weight compounds that can be targeted to the active site of virus proteinases without affecting cellular functions. Over the past years, both rational drug design based on X-ray crystallography and high-throughput screening of compound libraries have aided in the development of such inhibitors. The success of this approach is impressively illustrated by the recent advances in the treatment of human immunodeficiency virus (HIV) infections using HIV proteinase inhibitors (for reviews see Flexner, 1998; Patick & Potts, 1998; Wlodawer & Vondrasek, 1998). In the case of the nidovirus proteinases, it can be expected that additional potential target structures will be identified as the elucidation of interactions between proteinases and RNA or protein components of the replication complex proceeds.

The authors are indebted to all colleagues with whom they collaborated in the study of nidoviruses and who contributed results and ideas reviewed in this paper. The current work of A.E.G. is supported by Federal funds from the National Cancer Institute, National Institutes of Health, under contract no. NO1-CO-56000. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government. The current work of J.Z. is supported by grants from the Deutsche Forschungsgemeinschaft (SI 357/2-1, 2-2).

References

Agranovsky, A. A. (1996). Principles of molecular organization, expression, and evolution of closteroviruses: over the barriers. *Advances in Virus Research* **47**, 119–158.

Allaire, M., Chernaia, M. M., Malcolm, B. A. & James, M. N. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* **369**, 72–76.

Allende, R., Lewis, T. L., Lu, Z., Rock, D. L., Kutish, G. F., Ali, A., Doster, A. R. & Osorio, F. A. (1999). North American and European porcine reproductive and respiratory syndrome viruses differ in non-structural protein coding regions. *Journal of General Virology* **80**, 307–315.

Babé, L. M. & Craik, C. S. (1997). Viral proteases: evolution of diverse structural motifs to optimize function. *Cell* **91**, 427–430.

Baker, S. C., Shieh, C. K., Soe, L. H., Chang, M. F., Vannier, D. M. & Lai, M. M. (1989). Identification of a domain required for autoproteolytic cleavage of murine coronavirus gene A polyprotein. *Journal of Virology* **63**, 3693–3699.

Baker, S. C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A. E., Koonin, E. V. & Lai, M. M. (1993). Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. *Journal of Virology* **67**, 6056–6063.

Bazan, J. F. & Fletterick, R. J. (1988). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proceedings of the National Academy of Sciences, USA* **85**, 7872–7876.

Bi, W., Piñón, J. D., Hughes, S., Bonilla, P. J., Holmes, K. V., Weiss, S. R. & Leibowitz, J. L. (1998). Localization of mouse hepatitis virus open reading frame 1A derived proteins. *Journal of Neurovirology* **4**, 594–605.

Blom, N., Hansen, J., Blaas, D. & Brunak, S. (1996). Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. *Protein Science* **5**, 2203–2216.

Bonilla, P. J., Gorbalenya, A. E. & Weiss, S. R. (1994). Mouse hepatitis virus strain A59 RNA polymerase gene ORF 1a: heterogeneity among MHV strains. *Virology* **198**, 736–740.

Bonilla, P. J., Hughes, S. A., Piñón, J. D. & Weiss, S. R. (1995). Characterization of the leader papain-like proteinase of MHV-A59: identification of a new in vitro cleavage site. *Virology* **209**, 489–497.

Bonilla, P. J., Hughes, S. A. & Weiss, S. R. (1997). Characterization of a second cleavage site and demonstration of activity in trans by the papain-like proteinase of the murine coronavirus mouse hepatitis virus strain A59. *Journal of Virology* **71**, 900–909.

Boursnell, M. E. G., Brown, T. D. K., Foulds, I. J., Green, P. F., Tomley, F. M. & Binns, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *Journal of General Virology* **68**, 57–77.

Bredenbeek, P. J., Pachuk, C. J., Noten, A. F., Charite, J., Luytjes, W., Weiss, S. R. & Spaan, W. J. (1990*a*). The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Research* **18**, 1825–1832.

Bredenbeek, P. J., Snijder, E. J., Noten, F. H., den Boon, J. A., Schaaper, W. M., Horzinek, M. C. & Spaan, W. J. (1990*b*). The polymerase gene of corona- and toroviruses: evidence for an evolutionary relationship. *Advances in Experimental Medicine and Biology* **276**, 307–316.

Brierley, I., Boursnell, M. E., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D. & Inglis, S. C. (1987). An efficient ribosomal frame-shifting

signal in the polymerase-encoding region of the coronavirus IBV. *EMBO Journal* **6**, 3779–3785.

Brierley, I., Jenner, A. J. & Inglis, S. C. (1992). Mutational analysis of the 'slippery-sequence' component of a coronavirus ribosomal frameshifting signal. *Journal of Molecular Biology* **227**, 463–479.

Carrington, J. C. & Dougherty, W. G. (1988). A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proceedings of the National Academy of Sciences, USA* **85**, 3391–3395.

Carson, M. (1997). Ribbons. *Methods in Enzymology* **277**, 493–505.

Cavanagh, D. (1995). The coronavirus surface glycoprotein. In *The Coronaviridae*, pp. 73–113. Edited by S. G. Siddell. New York & London: Plenum Press.

Cavanagh, D. (1997). Nidovirales: a new order comprising Coronaviridae and Arteriviridae. *Archives of Virology* **142**, 629–633.

Choi, H.-K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G. & Wengler, G. (1991). Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature* **354**, 37–43.

den Boon, J. A., Snijder, E. J., Chirnside, E. D., de Vries, A. A., Horzinek, M. C. & Spaan, W. J. (1991). Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *Journal of Virology* **65**, 2910–2920.

den Boon, J. A., Faaberg, K. S., Meulenberg, J. J., Wassenaar, A. L., Plagemann, P. G., Gorbalenya, A. E. & Snijder, E. J. (1995). Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papainlike cysteine proteases. *Journal of Virology* **69**, 4500–4505.

Denison, M. R. & Perlman, S. (1986). Translation and processing of mouse hepatitis virus virion RNA in a cell-free system. *Journal of Virology* **60**, 12–18.

Denison, M. & Perlman, S. (1987). Identification of putative polymerase gene product in cells infected with murine coronavirus A59. *Virology* **157**, 565–568.

Denison, M. R., Zoltick, P. W., Hughes, S. A., Giangreco, B., Olson, A. L., Perlman, S., Leibowitz, J. L. & Weiss, S. R. (1992). Intracellular processing of the N-terminal ORF 1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events. *Virology* **189**, 274–284.

Denison, M. R., Hughes, S. A. & Weiss, S. R. (1995). Identification and characterization of a 65-kDa protein processed from the gene 1 polyprotein of the murine coronavirus MHV-A59. *Virology* **207**, 316–320.

Denison, M. R., Spaan, W. J., van der Meer, Y., Gibson, C. A., Sims, A. C., Prentice, E. & Lu, X. T. (1999). The putative helicase of the coronavirus mouse hepatitis virus is processed from the replicase gene polyprotein and localizes in complexes that are active in viral RNA synthesis. *Journal of Virology* **73**, 6862–6871.

de Vries, A. A. F., Horzinek, M. C., Rottier, P. J. M. & de Groot, R. J. (1997). The genome organization of the nidovirales: similarities and differences between arteri-, toro-, and coronaviruses. *Seminars in Virology* **8**, 33–47.

Dolja, V. V., Karasev, A. V. & Koonin, E. V. (1994). Molecular biology and evolution of closteroviruses: sophisticated build-up of large RNA genomes. *Annual Review of Phytopathology* **32**, 261–285.

Dong, S. & Baker, S. C. (1994). Determinants of the p28 cleavage site recognized by the first papain-like cysteine proteinase of murine coronavirus. *Virology* **204**, 541–549.

Dougherty, W. G. & Semler, B. L. (1993). Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiological Reviews* **57**, 781–822.

Eleouet, J. F., Rasschaert, D., Lambert, P., Levy, L., Vende, P. & Laude, H. (1995). Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus. *Virology* **206**, 817–822.

Flexner, C. (1998). HIV-protease inhibitors. *New England Journal of Medicine* **338**, 1281–1292.

Gao, H. Q., Schiller, J. J. & Baker, S. C. (1996). Identification of the polymerase polyprotein products p72 and p65 of the murine coronavirus MHV-JHM. *Virus Research* **45**, 101–109.

Gibson, C. A. & Denison, M. R. (1998). Coronavirus picornain-like cysteine proteinase. In *Handbook of Proteolytic Enzymes*, pp. 726–729. Edited by A. J. Barrett, N. D. Rawlings & J. F. Woessner. San Diego: Academic Press.

Godeny, E. K., Chen, L., Kumar, S. N., Methven, S. L., Koonin, E. V. & Brinton, M. A. (1993). Complete genomic sequence and phylogenetic analysis of the lactate dehydrogenase-elevating virus (LDV). *Virology* **194**, 585–596.

Goldbach, R. & Wellink, J. (1988). Evolution of plus-strand RNA viruses. *Intervirology* **29**, 260–267.

Gorbalenya, A. E. & Koonin, E. V. (1993). Comparative analysis of the amino acid sequences of the key enzymes of the replication and expression of positive-strand RNA viruses. Validity of the approach and functional and evolutionary implications. *Soviet Scientific Reviews Section D Physicochemical Biology Reviews* **11**, 1–84.

Gorbalenya, A. E. & Snijder, E. J. (1996). Viral cysteine proteinases. *Perspectives in Drug Discovery and Design* **6**, 64–86.

Gorbalenya, A. E., Koonin, E. V., Blinov, V. M. & Donchenko, A. P. (1988). Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Letters* **236**, 287–290.

Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M. & Koonin, E. V. (1989*a*). Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS Letters* **243**, 103–114.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989*b*). Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Research* **17**, 4847–4861.

Gorbalenya, A. E., Koonin, E. V. & Lai, M. M. (1991). Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. *FEBS Letters* **288**, 201–205.

Gorbalenya, A. E., den Boon, J. A. & Snijder, E. J. (1998*a*). Arterivirus papain-like cysteine endopeptidase α. In *Handbook of Proteolytic Enzymes*, pp. 693–695. Edited by A. J. Barrett, N. D. Rawlings & J. F. Woessner. San Diego: Academic Press.

Gorbalenya, A. E., Wassenaar, A. L. & Snijder, E. J. (1998*b*). Arterivirus Nsp2 cysteine endopeptidase. In *Handbook of Proteolytic Enzymes*, pp. 698–700. Edited by A. J. Barrett, N. D. Rawlings & J. F. Woessner. San Diego: Academic Press.

Grötzinger, C., Heusipp, G., Ziebuhr, J., Harms, U., Süss, J. & Siddell, S. G. (1996). Characterization of a 105-kDa polypeptide encoded in gene 1 of the human coronavirus HCV 229E. *Virology* **222**, 227–235.

Guarné, A., Tormo, J., Kirchweger, R., Pfistermueller, D., Fita, I. & Skern, T. (1998). Structure of the foot-and-mouth disease virus leader

protease: a papain-like fold adapted for self-processing and eIF4G recognition. *EMBO Journal* **17**, 7469–7479.

Herold, J., Raabe, T., Schelle-Prinz, B. & Siddell, S. G. (1993). Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology* **195**, 680–691.

Herold, J., Siddell, S. & Ziebuhr, J. (1996). Characterization of coronavirus RNA polymerase gene products. *Methods in Enzymology* **275**, 68–89.

Herold, J., Gorbalenya, A. E., Thiel, V., Schelle, B. & Siddell, S. G. (1998). Proteolytic processing at the amino terminus of human coronavirus 229E gene 1-encoded polyproteins: identification of a papain-like proteinase and its substrate. *Journal of Virology* **72**, 910–918.

Herold, J., Siddell, S. G. & Gorbalenya, A. E. (1999). A human RNA viral cysteine proteinase that depends upon a unique Zn^{2+}-binding finger connecting the two domains of a papain-like fold. *Journal of Biological Chemistry* **274**, 14918–14925.

Heusipp, G., Grötzinger, C., Herold, J., Siddell, S. G. & Ziebuhr, J. (1997 *a*). Identification and subcellular localization of a 41 kDa, polyprotein 1ab processing product in human coronavirus 229E-infected cells. *Journal of General Virology* **78**, 2789–2794.

Heusipp, G., Harms, U., Siddell, S. G. & Ziebuhr, J. (1997 *b*). Identification of an ATPase activity associated with a 71-kilodalton polypeptide encoded in gene 1 of the human coronavirus 229E. *Journal of Virology* **71**, 5631–5634.

Hughes, S. A., Bonilla, P. J. & Weiss, S. R. (1995). Identification of the murine coronavirus p28 cleavage site. *Journal of Virology* **69**, 809–813.

Johnston, S. C., Larsen, C. N., Cook, W. J., Wilkinson, K. D. & Hill, C. P. (1997). Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1·8 Å resolution. *EMBO Journal* **16**, 3787–3796.

Kräusslich, H. G. & Wimmer, E. (1988). Viral proteinases. *Annual Review of Biochemistry* **57**, 701–754.

Lai, M. M. & Cavanagh, D. (1997). The molecular biology of coronaviruses. *Advances in Virus Research* **48**, 1–100.

Lai, M. M., Baric, R. S., Brayton, P. R. & Stohlman, S. A. (1984). Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proceedings of the National Academy of Sciences, USA* **81**, 3626–3630.

Lawson, T. G., Gronros, D. L., Werner, J. A., Wey, A. C., DiGeorge, A. M., Lockhart, J. L., Wilson, J. W. & Wintrode, P. L. (1994). The encephalomyocarditis virus 3C protease is a substrate for the ubiquitin-mediated proteolytic system. *Journal of Biological Chemistry* **269**, 28429–28435.

Lawson, T. G., Gronros, D. L., Evans, P. E., Bastien, M. C., Michalewich, K. M., Clark, J. K., Edmonds, J. H., Graber, K. H., Werner, J. A., Lurvey, B. A. & Cate, J. M. (1999). Identification and characterization of a protein destruction signal in the encephalomyocarditis virus 3C protease. *Journal of Biological Chemistry* **274**, 9871–9880.

Lee, H. J., Shieh, C. K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhadzhyan, A. & Lai, M. M. (1991). The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology* **180**, 567–582.

Lim, K. P. & Liu, D. X. (1998). Characterization of the two overlapping papain-like proteinase domains encoded in gene 1 of the coronavirus infectious bronchitis virus and determination of the C-terminal cleavage site of an 87-kDa protein. *Virology* **245**, 303–312.

Liu, D. X. & Brown, T. D. (1995). Characterisation and mutational analysis of an ORF 1a-encoding proteinase domain responsible for proteolytic processing of the infectious bronchitis virus 1a/1b polyprotein. *Virology* **209**, 420–427.

Liu, D. X., Brierley, I., Tibbles, K. W. & Brown, T. D. (1994). A 100-kilodalton polypeptide encoded by open reading frame (ORF) 1b of the coronavirus infectious bronchitis virus is processed by ORF 1a products. *Journal of Virology* **68**, 5772–5780.

Lui, D. X., Tibbles, K. W., Cavanagh, D., Brown, T. D. & Brierley, I. (1995). Identification, expression, and processing of an 87 kDa polypeptide encoded by ORF1a of the coronavirus infectious bronchitis virus. *Virology* **208**, 48–57.

Liu, D. X., Xu, H. Y. & Brown, T. D. (1997). Proteolytic processing of the coronavirus infectious bronchitis virus 1a polyprotein: identification of a 10-kilodalton polypeptide and determination of its cleavage sites. *Journal of Virology* **71**, 1814–1820.

Liu, D. X., Shen, S., Xu, H. Y. & Wang, S. F. (1998). Proteolytic mapping of the coronavirus infectious bronchitis virus 1b polyprotein: evidence for the presence of four cleavage sites of the 3C-like proteinase and identification of two novel cleavage products. *Virology* **246**, 288–297.

Lu, Y. & Denison, M. R. (1997). Determinants of mouse hepatitis virus 3C-like proteinase activity. *Virology* **230**, 335–342.

Lu, Y., Lu, X. & Denison, M. R. (1995). Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. *Journal of Virology* **69**, 3554–3559.

Lu, X., Lu, Y. & Denison, M. R. (1996). Intracellular and in vitro-translated 27-kDa proteins contain the 3C-like proteinase activity of the coronavirus MHV-A59. *Virology* **222**, 375–382.

Lu, X. T., Sims, A. C. & Denison, M. R. (1998). Mouse hepatitis virus 3C-like protease cleaves a 22-kilodalton protein from the open reading frame 1a polyprotein in virus-infected cells and in vitro. *Journal of Virology* **72**, 2265–2271.

Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L. and others (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* **77**, 761–771.

Mayo, M. A. & Pringle, C. R. (1998). Virus taxonomy – 1997. *Journal of General Virology* **79**, 649–657.

Meulenberg, J. J., Hulst, M. M., de Meijer, E. J., Moonen, P. L., den Besten, A., de Kluyver, E. P., Wensvoort, G. & Moormann, R. J. (1993). Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV. *Virology* **192**, 62–72.

Meulenberg, J. J., Bos de Ruijter, J. N., van de Graaf, R., Wensvoort, G. & Moormann, R. J. (1998). Infectious transcripts from cloned genome-length cDNA of porcine reproductive and respiratory syndrome virus. *Journal of Virology* **72**, 380–387.

Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S. & James, M. N. (1997). Refined X-ray crystallographic structure of the poliovirus 3C gene product. *Journal of Molecular Biology* **273**, 1032–1047.

Nam, S. H., Copeland, T. D., Hatanaka, M. & Oroszlan, S. (1993). Characterization of ribosomal frameshifting for expression of pol gene products of human T-cell leukemia virus type I. *Journal of Virology* **67**, 196–203.

Nelsen, C. J., Murtaugh, M. P. & Faaberg, K. S. (1999). Porcine reproductive and respiratory syndrome virus comparison: divergent evolution on two continents. *Journal of Virology* **73**, 270–280.

Ng, L. F. & Liu, D. X. (1998). Identification of a 24-kDa polypeptide processed from the coronavirus infectious bronchitis virus 1a polyprotein by the 3C-like proteinase and determination of its cleavage sites. *Virology* **243**, 388–395.

Palmer, G. A., Kuo, L., Chen, Z., Faaberg, K. S. & Plagemann, P. G. (1995). Sequence of the genome of lactate dehydrogenase-elevating virus: heterogenicity between strains P and C. *Virology* **209**, 637–642.

Parks, G. D., Baker, J. C. & Palmenberg, A. C. (1989). Proteolytic cleavage of encephalomyocarditis virus capsid region substrates by precursors to the 3C enzyme. *Journal of Virology* **63**, 1054–1058.

Patick, A. K. & Potts, K. E. (1998). Protease inhibitors as antiviral agents. *Clinical Microbiological Reviews* **11**, 614–627.

Pedersen, K. W., van der Meer, Y., Roos, N. & Snijder, E. J. (1999). Open reading frame 1a-encoded subunits of the arterivirus replicase induce endoplasmic reticulum-derived double-membrane vesicles which carry the viral replication complex. *Journal of Virology* **73**, 2016–2026.

Piñón, J. D., Mayreddy, R. R., Turner, J. D., Khan, F. S., Bonilla, P. J. & Weiss, S. R. (1997). Efficient autoproteolytic processing of the MHV-A59 3C-like proteinase from the flanking hydrophobic domains requires membranes. *Virology* **230**, 309–322.

Piñón, J. D., Teng, H. & Weiss, S. R. (1999). Further requirements for cleavage by the murine coronavirus 3C-like proteinase: identification of a cleavage site within ORF1b. *Virology* **263**, 471–484.

Rueckert, R. R. & Wimmer, E. (1984). Systematic nomenclature of picornavirus proteins. *Journal of Virology* **50**, 957–959.

Ryan, M. D. & Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *Journal of General Virology* **78**, 699–723.

Schiller, J. J. & Baker, S. C. (1998). Coronavirus papain-like endopeptidases. In *Handbook of Proteolytic Enzymes*, pp. 681–683. Edited by A. J. Barrett, N. D. Rawlings & J. F. Woessner. San Diego: Academic Press.

Schiller, J. J., Kanjanahaluethai, A. & Baker, S. C. (1998). Processing of the coronavirus MHV-JHM polymerase polyprotein: identification of precursors and proteolytic products spanning 400 kilodaltons of ORF1a. *Virology* **242**, 288–302.

Schneider, T. D. & Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research* **18**, 6097–6100.

Seybert, A., Ziebuhr, J. & Siddell, S. G. (1997). Expression and characterization of a recombinant murine coronavirus 3C-like proteinase. *Journal of General Virology* **78**, 71–75.

Shi, S. T., Schiller, J. J., Kanjanahaluethai, A., Baker, S. C., Oh, J. W. & Lai, M. M. (1999). Colocalization and membrane association of murine hepatitis virus gene 1 products and de novo-synthesized viral RNA in infected cells. *Journal of Virology* **73**, 5957–5969.

Snijder, E. J. & Horzinek, M. C. (1993). Toroviruses: replication, evolution and comparison with other members of the coronavirus-like superfamily. *Journal of General Virology* **74**, 2305–2316.

Snijder, E. J. & Meulenberg, J. J. M. (1998). The molecular biology of arteriviruses. *Journal of General Virology* **79**, 961–979.

Snijder, E. J. & Spaan, W. J. M. (1995). The coronaviruslike superfamily. In *The Coronaviridae*, pp. 239–255. Edited by S. G. Siddell. New York & London: Plenum Press.

Snijder, E. J., den Boon, J. A., Bredenbeek, P. J., Horzinek, M. C., Rijnbrand, R. & Spaan, W. J. (1990). The carboxyl-terminal part of the putative Berne virus polymerase is expressed by ribosomal frameshifting and contains sequence motifs which indicate that toro- and coronaviruses are evolutionarily related. *Nucleic Acids Research* **18**, 4535–4542.

Snijder, E. J., Wassenaar, A. L. & Spaan, W. J. (1992). The 5′ end of the equine arteritis virus replicase gene encodes a papainlike cysteine protease. *Journal of Virology* **66**, 7040–7048.

Snijder, E. J., Wassenaar, A. L. & Spaan, W. J. (1994). Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. *Journal of Virology* **68**, 5755–5764.

Snijder, E. J., Wassenaar, A. L., Spaan, W. J. & Gorbalenya, A. E. (1995). The arterivirus Nsp2 protease. An unusual cysteine protease with primary structure similarities to both papain-like and chymotrypsin-like proteases. *Journal of Biological Chemistry* **270**, 16671–16676.

Snijder, E. J., Wassenaar, A. L., van Dinten, L. C., Spaan, W. J. & Gorbalenya, A. E. (1996). The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. *Journal of Biological Chemistry* **271**, 4864–4871.

Snijder, E. J., Wassenaar, A. L. & Gorbalenya, A. E. (1998*a*). Arterivirus papain-like cysteine endopeptidase *β*. In *Handbook of Proteolytic Enzymes*, pp. 695–697. Edited by A. J. Barrett, N. D. Rawlings & J. F. Woessner. San Diego: Academic Press.

Snijder, E. J., Wassenaar, A. L. & Gorbalenya, A. E. (1998*b*). Arterivirus serine endopeptidase. In *Handbook of Proteolytic Enzymes*, pp. 281–283. Edited by A. J. Barrett, N. D. Rawlings & J. F. Woessner. San Diego: Academic Press.

Soe, L. H., Shieh, C. K., Baker, S. C., Chang, M. F. & Lai, M. M. (1987). Sequence and translation of the murine coronavirus 5′-end genomic RNA reveals the N-terminal structure of the putative RNA polymerase. *Journal of Virology* **61**, 3968–3976.

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B. A. & Siddell, S. G. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO Journal* **2**, 1839–1844.

Strauss, J. H. & Strauss, E. G. (1988). Evolution of RNA viruses. *Annual Review of Microbiology* **42**, 657–683.

Strauss, J. H. & Strauss, E. G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiological Reviews* **58**, 491–562.

Teng, H., Piñón, J. D. & Weiss, S. R. (1999). Expression of murine coronavirus recombinant papain-like proteinase: efficient cleavage is dependent on the lengths of both the substrate and the proteinase polypeptides. *Journal of Virology* **73**, 2658–2666.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Computer Applications in the Biosciences* **10**, 19–29.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882.

Tibbles, K. W., Brierley, I., Cavanagh, D. & Brown, T. D. K. (1995). A region of the coronavirus infectious bronchitis virus 1a polyprotein encoding the 3C-like protease domain is subject to rapid turnover when expressed in rabbit reticulocyte lysate. *Journal of General Virology* **76**, 3059–3070.

Tibbles, K. W., Brierley, I., Cavanagh, D. & Brown, T. D. (1996). Characterization in vitro of an autocatalytic processing activity associated with the predicted 3C-like proteinase domain of the coronavirus avian infectious bronchitis virus. *Journal of Virology* **70**, 1923–1930.

Tibbles, K. W., Cavanagh, D. & Brown, T. D. (1999). Activity of a purified His-tagged 3C-like proteinase from the coronavirus infectious bronchitis virus. *Virus Research* **60**, 137–145.

van der Meer, Y., van Tol, H., Krijnse Locker, J. & Snijder, E. J. (1998). ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex. *Journal of Virology* **72**, 6689–6698.

van der Meer, Y., Snijder, E. J., Dobbe, J. C., Schleich, S., Denison, M. R., Spaan, W. J. & Krijnse Locker, J. (1999). Localization of mouse hepatitis virus nonstructural proteins and RNA synthesis indicates a role for late endosomes in viral replication. *Journal of Virology* **73**, 7641–7657.

van Dinten, L. C., Wassenaar, A. L., Gorbalenya, A. E., Spaan, W. J. & Snijder, E. J. (1996). Processing of the equine arteritis virus replicase ORF1b protein: identification of cleavage products containing the putative viral polymerase and helicase domains. *Journal of Virology* **70**, 6625–6633.

van Dinten, L. C., den Boon, J. A., Wassenaar, A. L., Spaan, W. J. & Snijder, E. J. (1997). An infectious arterivirus cDNA clone: identification of a replicase point mutation that abolishes discontinuous mRNA transcription. *Proceedings of the National Academy of Sciences, USA* **94**, 991–996.

van Dinten, L. C., Rensen, S., Gorbalenya, A. E. & Snijder, E. J. (1999). Proteolytic processing of the open reading frame 1b-encoded part of arterivirus replicase is mediated by nsp4 serine protease and is essential for virus replication. *Journal of Virology* **73**, 2027–2037.

van Marle, G., Dobbe, J. C., Gultyaev, A. P., Luytjes, W., Spaan, W. J. & Snijder, E. J. (1999*a*). Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. *Proceedings of the National Academy of Sciences, USA* **96**, 12056–12061.

van Marle, G., van Dinten, L. C., Spaan, W. J., Luytjes, W. & Snijder, E. J. (1999*b*). Characterization of an equine arteritis virus replicase mutant defective in subgenomic mRNA synthesis. *Journal of Virology* **73**, 5274–5281.

Wassenaar, A. L., Spaan, W. J., Gorbalenya, A. E. & Snijder, E. J. (1997). Alternative proteolytic processing of the arterivirus replicase ORF1a polyprotein: evidence that NSP2 acts as a cofactor for the NSP4 serine protease. *Journal of Virology* **71**, 9313–9322.

Wlodawer, A. & Vondrasek, J. (1998). Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annual Review of Biophysics and Biomolecular Structure* **27**, 249–284.

Yoo, D., Parker, M. D., Cox, G. J. & Babiuk, L. A. (1995). Zinc-binding of the cysteine-rich domain encoded in the open reading frame of 1B of the RNA polymerase gene of coronavirus. *Advances in Experimental Medicine and Biology* **380**, 437–442.

Ziebuhr, J. & Siddell, S. G. (1999). Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab. *Journal of Virology* **73**, 177–185.

Ziebuhr, J., Herold, J. & Siddell, S. G. (1995). Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *Journal of Virology* **69**, 4331–4338.

Ziebuhr, J., Heusipp, G. & Siddell, S. G. (1997). Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *Journal of Virology* **71**, 3992–3997.

Ziebuhr, J., Heusipp, G., Seybert, A. & Siddell, S. G. (1998). Substrate specificity of the human coronavirus 229E 3C-like proteinase. *Advances in Experimental Medicine and Biology* **440**, 115–120.